UNIVERSIDADE NOVA DE LISBOA

FACULDADE DE CIÊNCIAS MÉDICAS

TRANSCRIPTIONAL REGULATION OF THE MANNOSYLTRANSFERASE-ENCODING GENE *PIGM* IN INHERITED GLYCOSYLPHOSPHATIDYLINOSITOL (GPI) DEFICIENCY

JOANA RODRIGUES SIMÕES DA COSTA

Tese para obtenção do grau de Doutor em Ciências da Vida na especialidade em Biologia Celular e Molecular

2014

UNIVERSIDADE NOVA DE LISBOA

FACULDADE DE CIÊNCIAS MÉDICAS

TRANSCRIPTIONAL REGULATION OF THE MANNOSYLTRANSFERASE-ENCODING GENE *PIGM* IN INHERITED GLYCOSYLPHOSPHATIDYLINOSITOL (GPI) DEFICIENCY

JOANA RODRIGUES SIMÕES DA COSTA

TESE ORIENTADA PELO DR. ANTÓNIO ALMEIDA E PELO PROFESSOR ANASTASIOS KARADIMITRIS

Tese para obtenção do grau de Doutor em Ciências da Vida na especialidade em Biologia Celular e Molecular

2014

Aos meus pais e à minha irmã

Acknowledgements

I would like to thank both my supervisors, Professor Tassos Karadimitris and Dr. António Almeida. To Tassos, for giving me the opportunity to carry out my PhD research in his lab. His scientific guidance and particularly his trust in me were fundamental to the development of my project and for helping me to believe in myself as a researcher. I will never forget his inspirational words: "TFs are the masters of the Universe!". To Dr. António Almeida for giving me the opportunity to undertake my PhD and for understanding and supporting my decisions on what was best for my research.
I also wish to thank Fundação para a Ciência e Tecnologia (FCT) for funding my PhD studies.

A special acknowledgement goes to Valentina Caputo. The number of words that I am allowed to write on this page are certainly not enough to express my appreciation for the person that helped me the most. Her ongoing support, motivation, sharing of knowledge, thoughts, ideas and friendship kept me going. All our discussion made me grow as a researcher and gave me belief in myself. Without her I couldn't have given any "colour" to my research or completed this work. She was truly my greatest support in the lab and an inspiration as a post-doc. Gracias Vale. I'll never forget..

I must acknowledge Professor Irene Roberts for her support. Special thanks also go to Dr. Ian Sudbery for his collaboration and assistance with the bioinformatics aspect of this project.

I would like to thank my friend and colleague Kelly, for helping me, even without knowing, to make the big decision to stay in London to carry out my research. Her joy and support in the lab were fundamental to my work and her confidence an inspiration.
For all her help and strong support I wish to thank Katerina. Special thanks go to David O'Connor, for his care, strong support and friendship (including sharing good music and the English lessons!). Thanks to Antonella for her help in the lab and enormous inspiration. Neha and Deena, thank you for your attention! To Luciana, for making each day in the lab fun and making sure things ran smoothly. To Maria..for listening me so many times…

I also would like to thank all the other members of the lab for their trust and help.

A very special acknowledgement goes to my friend and flatmate Cristina, for her support, for her care and joy and for always being there. Thanks to Rute and Sérgio for listening and caring and for always being so supportive. To Elena and Gigi, for their help and support. A special acknowledgement goes to Ilaria Marigo for understanding me so well and for all the advice.

Gostaria de agradecer à Jacinta por toda a ajuda científica e pessoal, incluindo os conselhos e palavras de ânimo e de apoio, desde o inicio até ao final do meu doutoramento. Obrigada Jace!
Ao Bruno e ao Hélio pelo apoio e à Gabriela por tudo o que me ensinou no laboratório.
À Sofia, pela inalterável amizade, pela capacidade de mesmo à distancia me ter ouvido e apoiado tantas vezes e por ter contribuído para que este desafio tivesse sido mais leve. À Cristiana (Flatinha), à Rita D. e ao Jaime, pelo apoio, pela amizade e por todos os momentos especiais de boa disposição. Às LHO, por estarem sempre presentes, quer nos bons quer nos maus momentos . Martinha, Johny, Lara e Ju. Obrigada. À Cheila e ao Bruno, pela ajuda e partilha. Obrigada também aos restantes elementos do CIPM. E às minhas amigas de faculdade, Telma, Débora e Teresa.

Gostaria de agradecer aos meus tios Nita e Quim por ouvirem os meus desabafos e ainda que nas minhas poucas palavras consigam sempre compreender-me tão bem. O vosso apoio durante o meu percurso tem sido fundamental.

E finalmente aos meus pais e à minha irmã. Por me terem encorajado sempre a seguir o meu caminho e apoiado em todas as decisões mesmo que isso tenha significado a minha ausência.. Pela confiança que depositaram em mim, por me ouvirem sempre que eu mais preciso e por me fazerem sorrir!
Quero agradecer especialmente à minha mãe por ser o meu maior suporte (e o de todos nós).
Este meu trabalho é dedicado a voçês.

List of Publications

Caputo VS, **Costa JR**, Makarona K, *et al*. Mechanism of Polycomb recruitment to CpG islands revealed by inherited-disease-associated mutation. *Human Molecular Genetics*. 2013, 22 (16), pp.3187-3194.

Makarona, K, Caputo VS, **Costa JR**, *et al*. Transcriptional and epigenetic basis for restoration of G6PD enzymatic activity in human G6PD-deficient cells. *Blood*. 2014, 124 (1), pp.134-141.

Manuscript in preparation: Cell type-specific transcriptional regulation of *PIGM* underpins the divergent cell and clinical phenotype in inherited glycosylphosphatidylinositol deficiency. **Costa JR**, Caputo VS, Makarona, K *et al.* 2014.

List of Figures and Tables

Figures

Tables

List of abbreviations

AA - Aplastic anaemia

ALP - Alkaline phosphatase

BFU-E - Burst forming unit-erythroid

bp - Base pair

BMF - Bone marrow failure

BRE - TFIIB recognition element

CFU-E - Colony-forming unit-erythroid

CGI - CpG island

ChIP - Chromatin immunoprecipitation

CLP - Common lymphoid progenitor

CMP - Common myeloid progenitor

DBD - DNA binding domain

DHSs - DNase I hypersensitive sites

DN - Dominant negative

DPE - Downstream core promoter element

ER - Endoplasmic reticulum

EPO - erythropoietin

Ery - Erythroblast

FLAER - Fluorescein-labeled proaerolysin

GPI - Glycosylphosphatidylinositol

GPI-APs - GPI -anchored proteins

GPI-ET - GPI-ethanolamine

GPI-MT - GPI -mannosyltransferase

GTFs - General transcription factors

GWAS - Genome-wide association study

HATS - Histone Acetyltransferases

HDACs - Histone Deacetylases

HPMR - Hyperphosphatasia mental retardation

HSC - Haematopoietic stem cell

IGD - Inherited GPI Deficiency

INR - Initiator

kd - Knock-down

KLF - Krüppel-like factor

LBCL - Lymphoblastoid B cell line

LCR - Locus control region

MEP - megakaryocyte/erythroid progenitor

min - Minute

MitA - Mithramycin A

MNase - Micrococcal nuclease

NaBu - Sodium butyrate

NDR - Nucleosome-depleted region

O/N - Overnight

PcG - Polycomb

PRC - Polycomb-repressive complex

PolIII - RNA Polymerase III

PNH - Paroxysmal nocturnal haemoglobinuria

RBC - Red blood cells

REAA - Restriction enzyme accessibility assay

RNA PolII - RNA Polymerase II

rpm - Rotations per minute

RT - Room temperature

SCF - Stem cell factor

SNP - Single nucleotide polymorphism

Sp - Specificity protein

TAD - Transactivation domain

TF - Transcription factor

TFBS - Transcription factor binding site

TBP - TATA-binding protein

TSS - Transcription start site

4C - Circular Chromosome Conformation Capture

Resumo

O glicosilfosfatidilinositol (GPI) é um complexo glicolipídico utlizado por dezenas de proteínas, o qual medeia a sua ancoragem à superfície da célula. Proteínas de superfície celular ancoradas a GPI apresentam várias funções essenciais para a manutenção celular. A deficiência na síntese de GPI é o que caracteriza principalmente a deficiência hereditária em GPI, um grupo de doenças autossómicas raras que resultam de mutações nos genes *PIGA*, *PIGL*, *PIGM*, *PIGV*, *PIGN*, *PIGO* e *PIGT*, os quais sao indispensáveis para a biossíntese do GPI.

Uma mutação pontual no motivo rico em GC -270 no promotor de *PIGM* impede a ligação do factor de transcrição (FT) Sp1 à sua sequência de reconhecimento, impondo a compactação da cromatina, associada à hipoacetilação de histonas, e consequentemente, impedindo a transcrição de *PIGM*. Desta forma, a adição da primeira manose ao GPI é comprometida, a síntese de GPI diminui assim como as proteínas ligadas a GPI à superficie das células. Pacientes com Deficiência Hereditária em GPI-associada a *PIGM* apresentam trombose e epilesia, e ausência de hemólise intravascular e anemia, sendo que estas duas últimas características definem a Hemoglobinúria Paroxística Nocturna (HPN), uma doença rara causada por mutações no gene *PIGA*.

Embora a mutação que causa IGD seja constitutiva e esteja presente em todos os tecidos, o grau de deficiência em GPI varia entre células do mesmo tecido e entre células de tecidos diferentes. Por exemplo nos granulócitos e linfócitos B a deficiência em GPI é muito acentuada mas nos linfócitos T, fibroblastos, plaquetas e eritrócitos é aproximadamente normal, daí a ausência de hemólise intravascular.

Os eventos transcricionais que estão na base da expressão diferencial da âncora GPI nas células hematopoiéticas são desconhecidos e constituem o objectivo geral desta tese.

Em primeiro lugar, os resultados demonstraram que os níveis de *PIGM* mRNA variam entre células primárias hematopoiéticas normais. Adicionalmente, a configuração dos nucleossomas no promotor de *PIGM* é mais compacta em células B do que em células eritróides e tal está correlacionado com os níveis de expressão de *PIGM*, isto é, inferior nas células B. A presença de vários motivos de ligação para o FT específico da linhagem megacariocítica-eritróide GATA-1 no promotor de *PIGM* sugeriu que GATA-1 desempenha um papel regulador na sua transcrição. Os resultados mostraram que muito possivelmente GATA-1 desempenha um papel repressor em vez de activador da expressão de *PIGM*. Resultados preliminares sugerem que KLF1, um factor de transcrição restritamente eritróide, regula a transcrição de *PIGM* independentemente do motivo -270GC.

Em segundo lugar, a investigação do papel dos FTs Sp demonstrou que Sp1 medeia directamente a transcrição de *PIGM* em ambas as células B e eritróide. Curiosamente, ao contrário do que acontece nas células B, em que a transcrição de *PIGM* requer a ligação do FT geral Sp1 ao motivo -270GC, nas células eritróides Sp1 regula a transcrição de *PIGM* ao ligar-se a montante e não ao motivo -270GC. Para além disso, demonstrou-se que Sp2 não é um regulador directo da transcrição de *PIGM* quer nas células B quer nas células eritróides.
Estes resultados explicam a ausência de hemólise intravascular nos doentes com IGD associada a *PIGM*, uma das principais características que define a HPN.

Por último, resultados preliminares mostraram que a repressão da transcrição de *PIGM* devida à mutação patogénica -270C>G está associada com a diminuição da frequência de interacções genómicas em *cis* entre *PIGM* e os seus genes "vizinhos", sugerindo adicionalmente que a regulação de *PIGM* e desses genes é partilhada.

No seu conjunto, os resultados apresentados nesta tese contribuem para o conhecimento do controlo transcricional de um gene *housekeeping*, específico-de-tecido, por meio de FTs genéricos e específicos de linhagem.

Abstract

Glycosylphosphatidylinositol (GPI) is a complex glycolipid used by dozens of proteins for cell surface anchoring. GPI-anchored proteins have various functions that are essential for the cellular maintenance. Defective GPI biosynthesis is the hallmark of inherited GPI deficiency (IGD), a group of rare autosomal diseases caused by mutations in *PIGA*, *PIGL*, *PIGM*, *PIGV*, *PIGN*, *PIGO* and *PIGT*, all genes indispensable for GPI biosynthesis.

A point mutation in the -270GC-rich box in the core promoter of *PIGM* disrupts binding of the transcription factor (TF) Sp1 to it, imposing nucleosome compaction associated with histone hypoacetylation, thus abrogating transcription of *PIGM*. As a consequence of *PIGM* transcriptional repression, addition of the first mannose residue onto the GPI core and thus GPI production are impaired; and expression of GPI-anchored proteins on the surface of cells is severely impaired. Patients with *PIGM*-associated IGD suffer from life-threatening thrombosis and epilepsy but not intravascular haemolysis and anaemia, two defining features of paroxysmal nocturnal haemoglobinuria (PNH), a rare disease caused by somatic mutations in *PIGA*.

Although the disease-causing mutation in IGD is constitutional and present in all tissues, the degree of GPI deficiency is variable and differs between cells of the same and of different tissues. Accordingly, GPI deficiency is severe in granulocytes and B cells but mild in T cells, fibroblasts, platelets and erythrocytes, hence the lack of intravascular haemolysis.

The transcriptional events underlying differential expression of GPI in the haematopoietic cells of *PIG-M*-associated IGD are not known and constitute the general aim of this thesis.

Firstly, I found that *PIGM* mRNA levels are variable amongst normal primary haematopoietic cells. In addition, the nucleosome configuration in the promoter of *PIGM* is more compacted in B cells than in erythroid cells and this correlated with the levels of *PIGM* mRNA expression, i.e., lower in B cells. The presence of several binding sites for GATA-1, a mega-erythroid lineage-specific transcription

factor (TF), at the *PIGM* promoter suggested that GATA-1 has a role on *PIGM* transcription. My results showed that GATA-1 in erythroid cells is most likely a repressor rather than an activator of *PIGM* expression. Preliminary data suggested that KLF1, an erythroid-specific TF, regulates *PIGM* transcription but independently of the -270GC motif.

Secondly, investigation of the role of the Sp TFs showed that Sp1 directly mediates *PIGM* transcriptional regulation in both B and erythroid cells. However, unlike in B cells in which active *PIGM* transcription requires binding of the generic TF Sp1 to the -270GC-rich box, in erythroid cells, Sp1 regulates *PIGM* transcription by binding upstream of but not to the -270GC-rich motif. Additionally, I showed that Sp2 is not a direct regulator of *PIGM* transcription in B and erythroid cells.

These findings explain lack of intravascular haemolysis in *PIGM*-associated IGD, a defining feature of PNH.

Lastly, preliminary work shows that transcriptional repression of *PIG-M* by the pathogenic -270C>G mutation is associated with reduced frequency of *in cis* genomic interactions between *PIGM* and its neighbouring genes, suggesting a shared regulatory link between these genes and *PIGM*.

Altogether, the results presented in this thesis provide novel insights into tissue-specific transcriptional control of a housekeeping gene by lineage-specific and generic TFs.

Chapter 1 – Introduction

1.1 Regulation of gene expression

Gene expression is the fundamental process by which the genetic information is converted to messenger RNA (mRNA) that is ultimately translated into functional proteins essential to the cell.

Control of gene expression has to be dynamic and highly regulated in order that spatial and temporal patterns that ensure cell and tissue development and function are achieved. For this purpose, gene expression is fine-tuned at different levels starting with transcription. RNA Polymerase II (RNA PolII) binding and active or repressed gene transcription is regulated by the integration of in *cis* DNA sequence information (promoters, enhancers and insulators), DNA elements with processes that determine chromatin packaging, histone and DNA modification (epigenetic marks). Next, I will provide an overview of transcriptional regulation of gene expression.

1.1.1 Transcriptional regulation of gene expression

Typically, protein-coding genes contain two distinct modules of *cis*-acting DNA elements: i) a promoter, which is composed by a core promoter and nearby (or proximal) promoter, and ii) distal regulatory elements, which can be enhancers, silencers, insulators or locus control regions (LCR) (Fig.1).

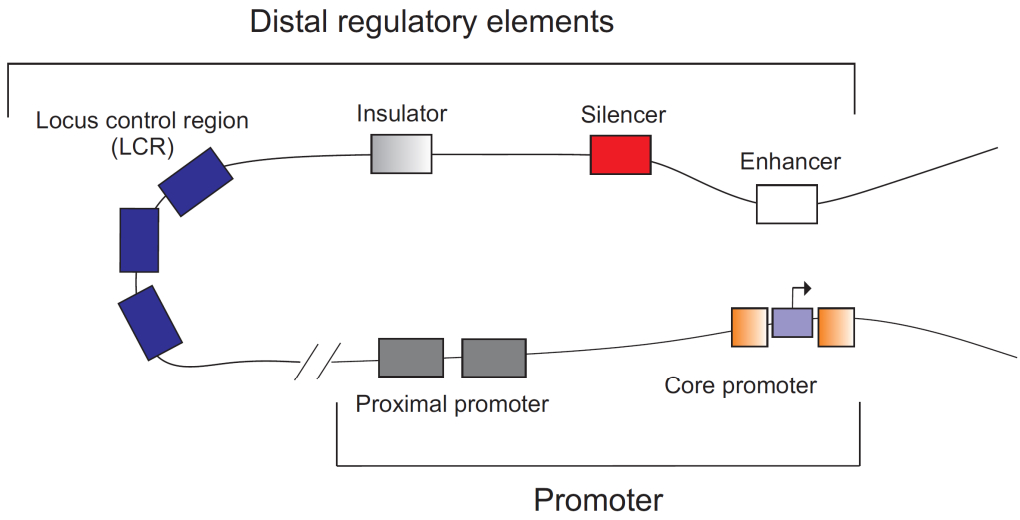


Fig. 1 - Schematic representation of a typical gene regulatory region (adapted from Maston, G.A. *et al*, 2006). The promoter region comprises the proximal and the core promoter and usually spans ≤ 1Kb. The core promoter normally encompasses the transcription start site (TSS), from where transcription is initiated (black arrow).

Whereas the core promoter directs the initiation of transcription at precise sites by nucleating the Pre-initiation complex (PIC) and the RNA PolII, the proximal promoter and distal regulatory elements can activate or repress transcription. Distal regulatory elements, like enhancers or LCR, dispersed over tens to thousands of kilobases from the transcription start site (TSS), regulate gene transcription by forming DNA loops, in which distal DNA-bound proteins are nucleated to the core promoter. A very well-studied example is the β-globin locus where the distal elements - deoxyribonuclease I (DNase I) hypersensitive sites (DHSs) – in the LCR physically interact with the β-globin gene, *HBB*, by chromosomal looping (1, 2). This mechanism of gene regulation has also been demonstrated in transitional states of α-globin expression during erythroid differentiation (3).

The method most used to study intrachromosomal interactions between *cis*-acting elements is called 3C (chromosome conformation capture) (4) and recently, methods derived from this technique have been developed (5, 6) and applied at the single cell level (7).

Alterations in DNA regulatory elements have been associated with human diseases and in fact, it is estimated that between 1 to 2% of disease-causing point mutations are in regulatory regions of the genome, the majority located within 1Kb of the TSS (8). Nevertheless, *cis*-acting mutations do not necessarily affect the nearby gene. In agreement, a genome-wide association study (GWAS) has recently linked single nucleotide polymorphisms (SNPs) within DHSs not only to adjacent genes but also to new distant gene targets. Interestingly, most SNPs within DHSs overlapped with a transcription factor recognition sequence further suggesting the disruption of common transcriptional networks in disease (9).

As reviewed by Butler J. *et al* (10), the core promoter is the minimal DNA sequence encompassing the TSS (extending approximately 35 nucleotides upstream and downstream), that is sufficient to direct initiation of transcription by "docking" the basal transcriptional machinery; the proximal promoter, immediately upstream of the core promoter, spans a few hundred nucleotides. Proximal promoters, as well as distal regulatory elements, contain short DNA sequences known to anchor specific transcription factors (TFs), proteins that bind DNA and modulate transcription. TFs can be subdivided into two groups: generic and cell-type specific. Generic TFs are ubiquitously expressed in all cell types whereas cell-type specific TFs have restricted expression.
The DNA sequences bound by TFs, called transcription factor binding sites (TFBS), are typically constituted by 6 to 12 nucleotides although binding specificity is usually determined by 4 to 6 positions within the site (11).

1.1.2 The transcriptional machinery

Protein-coding genes are transcribed by RNA PolII which generates mRNA transcripts from a template DNA. Accurate and efficient transcription initiation involves various factors which can be classified in three groups: general transcription factors (GTFs) that include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH; activators and co-activators. GTFs assemble in the core promoter in an ordered manner together with RNA PolII at the TSS. Transcriptional activity

derived from binding of the general factors and RNA PolII at the core promoter - basal transcription -, is usually low but can be potentiated by transcription factors that once recruited stabilize the transcriptional machinery. Accordingly, transcriptional activity is greatly stimulated or repressed by co-activators or co-repressors, respectively, localised upstream at the core promoter. Co-factors lack DNA binding properties and therefore interact physically with the DNA through bound TFs. For instance, Mediator, a multiprotein complex, is a co-activator known to target RNA PolII and recently, various crucial aspects of transcriptional regulation have been assigned to it (12). RNA PolII core promoters are functionally and structurally diverse. A small proportion contains a TATA box located 30 base pairs (bp) upstream of the TSS that guides RNA PolII to initiate transcription ("sharp" promoters). Usually, TATA promoters exhibit a low GC dinucleotide content and are associated with tissue-specific transcription. On the other hand, the majority of the promoters contain long (on average 1000bp) GC-rich DNA sequences called CpG island (CGIs) and correlate with the presence of multiple TSSs ("broad" promoters) (13-15). CGIs-associated promoters have been linked for long to housekeeping (ubiquitously expressed) genes, but is now clear that tissue-specific genes also contain CGI promoters (16, 17). Interestingly, some promoters contain both a TATA box and a CpG island, like in the erythropoietin gene (18, 19) and new studies have suggested that promoters of this type are capable of both TATA-dependent and TATA-independent transcriptional initiation (20). Therefore, the initial distinction between high and low-CpG promoters tends to fall in disuse and as suggested by Lenhard, B *et al*, 2012, (21) dividing promoters into "sharp" and "broad" may provide a better functional classification.

Apart from the TATA box, core promoters also contain other functional elements including initiator (INR), TFIIB recognition element (BRE), downstream core promoter element (DPE), amongst others (10).

Defining the boundaries of core promoters is challenging due to their complexity. Most genes use more than one core promoter (alternative promoter usage), typically located hundreds or thousands of nucleotides apart, and most promoters contain a cluster of TSSs rather than a single, well defined TSS. The use of alternative promoters entails the generation of at least two different mRNA

isoforms that might generate distinct protein products and moreover, it is associated with patterns of gene expression, i.e., ubiquitous or tissue-specific/developmental stage-specific (22). There are examples in the literature in which the generation of a new promoter element causes disease by interfering with normal gene expression (23).

1.1.3 Epigenetic control of transcription

TFs usually are part of multiprotein complexes. Beyond recruiting co-activators and co-repressors, TFs also recruit histone-modifying enzymes and remodelling complexes that modify the nucleosomes, the basic unit of chromatin, in a dynamic manner. The term **Epigenetics** refers transferable regulatory information that controls gene transcription by processes other than changes in the underlying DNA sequence.

The nucleosome, the basic unit of chromatin, is composed by an octamer of four core histones H2A, H2B, H3 and H4, wrapped around a string of 147 base pairs of DNA. Consecutive nucleosomes are regularly arranged and connected by the DNA linker histone H1 and once compacted in higher-order structures originate the chromosome (Fig. 2).

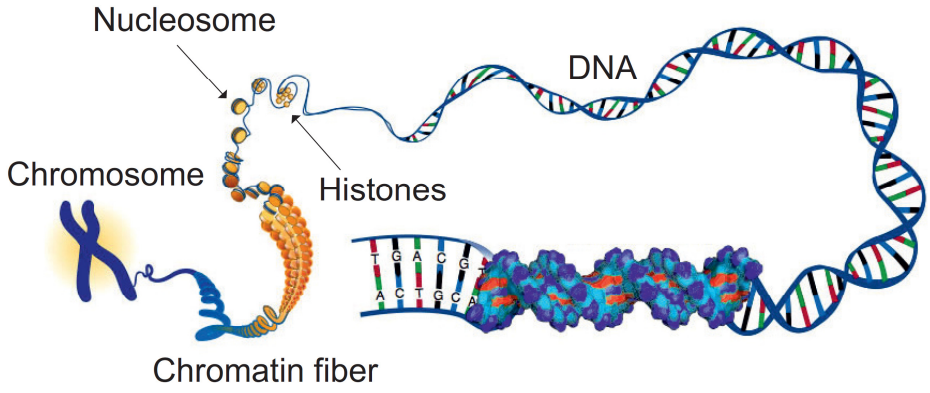


Fig. 2 – Levels of structural organization of the DNA and histone proteins. DNA wrapped around a globular structure of the histone octamer constitutes the nucleosome. The chromatin fiber consists in the higher order of chromatin organisation; further

packaging of the chromatin fiber originates the chromosome. (Adapted from image on the National Human Genome Research Institute website; author - Darryl Leja).

In the genome, the type and positioning of the nucleosomes is determined by a variety of factors including, DNA sequence, activators, nucleosome remodelers and the RNA PolII transcription machinery (24).

The histones that constitute the nucleosomal unit are proteins with a globular domain and an N-terminal "tail". Whilst the globular domains interact with each other in order to form the octameric core, the histone "tails" undergo post-translational modifications (PTM) at specific aminoacid residues. Those modifications, mediated by histone-modifying enzymes, include acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deamination and proline isomerization. Acetylation of histone "tails" is one of the best studied modifications. Usually, acetylation of histone "tails" by histone acetyltransferases (HATs) is associated with active gene expression whereas hypoacetylation at specific promoters by histone deacetylases (HDACs) is associated with repressed genes (25). Nevertheless, HDACs have been also found at active genes suggesting that gene expression results from a dynamic, balanced process between acetylation and deacetylation (26).

Another epigenetic mark that modulates gene expression is DNA methylation. DNA methylation occurs at the cytosine nucleotides that constitute the CpG modules and it is typically associated with gene silencing (27).

1.1.4 Transcription factors and DNA-binding specificity

It is estimated that there are approximately 1,400 transcription factors containing sequence specific DNA-binding properties (28).

The DNA binding specificity of a TF is primary defined by the **DNA binding domain (DBD)** and according to it, TFs can be classified in different families: cysteine-rich zinc finger, homeobox, helix-loop-helix (HLH), basic leucine zipper (bZIP), forkhead, ETS, Pit-Oct-Unc (POU) (11). In addition to the DBD, TFs also

contain a **transactivation domain (TAD)** which by recruiting the other elements of the transcriptional machinery stimulates transcription.

Whereas the DBD of TFs has been studied *in vitro* for many decades by techniques such as Electrophoretic Mobility Shift Assay (EMSA), full-length transcription factor occupancy has been studied *in vivo* only in the past few years by methods including Chromatin Immunoprecitipation (ChIP) and its derivatives (ChIP-chip, ChIP-seq and ChIP-exo) and DamID. DNase I hypersensitivity and Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) have been also extensively used to investigate chromatin structure. The former is based on the principle that TF binding to the DNA generates a remodelled chromatin state more permissive to nuclease digestion. Studies have recently suggested the use of DHSs profile as quantitative benchmark in lineage specification (29-31).

Across the genome, only a small fraction of predicted consensus motifs is occupied by TFs (32, 33) suggesting that the presence of a consensus motif is not always sufficient to explain TF binding. Understanding what determines TF occupancy at particular consensus motifs is a very interesting and emerging area of research in the field of regulation of gene expression.
The sequence flanking the core motif (34), cofactor binding (35) and the chromatin landscape (DHSs and histone marks) (36, 37) are known to influence TF binding and consequently gene expression. Additionally, spacing and orientation between dimeric sites also influence TF binding specificity (34).
Conversely, TF binding also influences the chromatin landscape (29, 38).

In what respects cell-type specific TF binding preferences, studies in the field are even less numerous. The contribution of DNase accessibility along with DNA sequence and histone modifications in cell type-specific TF binding was recently investigated genome-wide by using a computational approach. DNase accessibility not always explains cell-type specific TF binding as in sites where binding sites were accessible in both erythroid and lymphoid cell lines, TFs, such as JUND and YY1, occupied their cognate motifs only in one cell type. Cell-type specific TF binding was therefore more reliable on the “preferential” sequence (determined by the presence of heterodimer motifs for example) rather than on

chromatin accessibility (39). Another study, using as a model the estrogen receptor (ER) compared, in multiple cell types, the properties of the cell-type specific and shared TF binding sites. Most shared sites contained high-affinity cognate sequences and did not show a dependency on open chromatin. On the other hand, cell specific sites harboured low-affinity sequence elements and were associated with chromatin accessibility (40). Other investigations suggest that the quantitative difference in TF-binding between cell types, rather than an ON/OFF event, should be taken into account when studying TF binding properties (41).

Together, these studies highlight the importance of the underlying DNA sequence in TF binding specificity. They reveal the complexity of TF binding and the importance in understanding DNA binding properties of individual TFs in a cell-type specific context in vivo.

Overall, the functional importance of differential and cell type-specific preferences of TF binding to regulatory DNA at a single gene level in health and in disease has not been determined. In my PhD project, I took advantage of a pathogenic promoter mutation that causes Inherited Glycosylphosphatidylinositol (GPI) Deficiency, i.e, IGD, to address these issues.

1.1.5 Deregulation of transcription in disease

Many diseases and syndromes have been linked to alterations in components of transcriptional regulation. Apart from mutations in regulatory elements, mutations in transcription factors, co-factors, chromatin regulators, and non-coding RNAs have been also reported.

Regarding TFs, alterations in their levels, structure or function impact changes in wide gene regulatory networks and not surprisingly, have implications in various human diseases, including cancer (42, 43), autoimmunity and inflammation (44, 45), cardiovascular disease (46) amongst others. Importantly, a vast number of TFs has been associated with developmental defects, revealing their importance during early stages of development (47-49). The capacity of some TFs in auto-regulate their own expression also highlights their importance in maintaining appropriate gene expression (50, 51).

The majority of the alterations in TFs relate by far with loss of function of the protein. Some others relate with generation of a fusion protein or de-regulated expression due to translocations, mutations in promoters or enhancers and very few are associated with gain of function or amplification (reviewed by Lee and Young, 2013 (52)).
Changes in gene regulatory regions normally alter transcription either by gaining or loss of a transcription factor binding site. For instance, a SNP in the α-globin cluster generates a new GATA-1 binding site, which by recruiting an erythroid-specific complex to the new promoter causes α-thalassemia (23). In the Pierre Robin syndrome, a heterozygous point mutation alters binding of the TF MSX1, abrogating the enhancer function, thus causing cleft palate (53).

One of the best studied Mendelian disorders caused by a mutation affecting TF binding is the *PIGM*-associated Inherited GPI Deficiency. It is caused by a hypomorphic C>G point mutation in the core promoter of the GPI mannosyltransferase I (GPI-MTI), PIGM, abrogating binding of the TF Sp1 (38).

In the subsequent sections, first, I will give an overview of the GPI biosynthetic pathway and the diseases associated with it followed by a more extensive description of the clinical, cellular, biochemical and 'transcriptional' phenotype of *PIGM*-associated GPI deficiency.

1.2 The Glycosylphosphatidylinositol (GPI) biosynthetic pathway

1.2.1 GPI-mediated anchoring of proteins in the cell membrane

Whereas a large number of proteins are embedded in the phospholipid bilayer of the cell membrane through a transmembrane arrangement, some surface proteins are anchored on the cell membrane through covalent linkage to GPI (54).
Glycosylphosphatidylinositol (GPI)-anchored proteins (GPI-APs) constitute a unique class of proteins associated with the cell membrane, intersecting only the

external monolayer of the cell membrane. The GPI anchor, a complex glycolipid consisting of a phosphatidylinositol (PI) moiety and a glycan core, can be attached to the C- terminus of dozens of proteins. GPI-APs have various functions in the cell, including enzymes, complement regulatory proteins, and signal transduction (summarised on Table 1).

Table 1 – Examples of GPI-APs and respective function in mammalian cells [adapted from (55)]

GPI-anchored protein	Function
Alkaline phosphatase **5'-nucleotidase** **Acetylcholinesterase** **Dipeptidase**	Enzymes
LFA-3 **NCAM**	Cell-cell interaction and adhesion
CD55 (DAF) **CD59**	Complement regulation
Thy-1 **CD14** **Carcinoembryonic antigen (CEA)** **CD52**	Mammalian antigens
CD16b **Folate-binding protein**	Others

The GPI backbone structure EtN-P-6Manα1-2Manα1-6Manα1-4GlcNα1-6*myo*Inositol-phospholipid (where EtN-P is ethanolamine phosphate; Man is mannose and GlcN is glucosamine) (Fig. 4) is shared among various eukaryotic GPI-APs, such as those in yeast (*Sacccharomyces cerevisiae*), protozoan parasites where they are highly abundant (*Trypanosoma brucei* and *Plasmodium falciparum*) and plants (*Pyrus communis*).

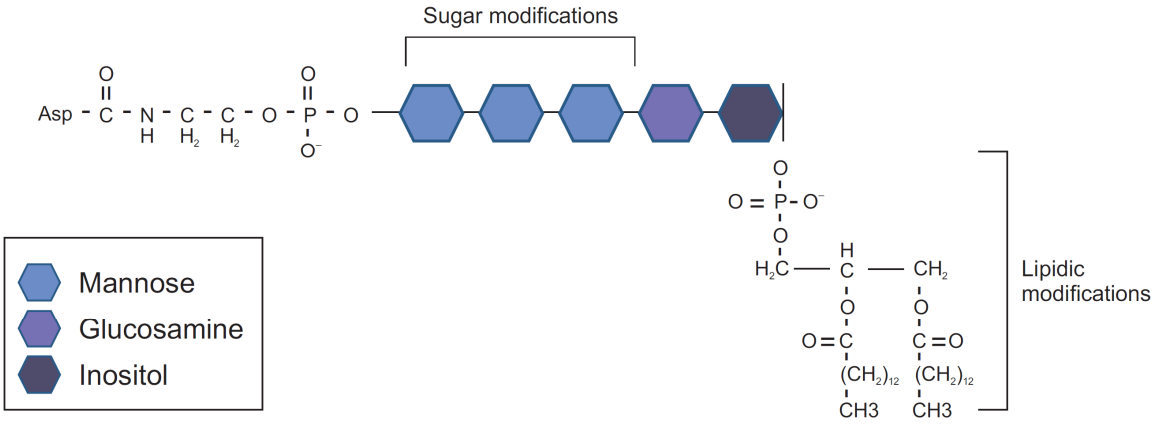


Fig. 3 – Structure of the GPI anchor common core. Adapted from Mayor and Riezman, 2004 (56).

Despite their common core, there is variability in both glycan and lipidic components of the GPI anchors, which might occur either before or after attachment to the proteins (56). Those variations are believed to have a biological significance, namely, in the lateral mobility of GPI-APs and in trafficking. However, the evolutionary purpose of this structurally diverse manner of expressing proteins, as well as the functional role of GPI itself, beyond the physical anchoring, remain unknown. The importance of GPI anchoring in mammals has been highlighted by the fact germ line disruption of GPI biosynthesis is embryonic lethal. For instance, *Pig-a* knock out mice presented neurodevelopmental defects and failed to develop beyond the ninth day of gestation (57).

The function of the enzymes involved in the biosynthesis of the GPI has been well characterised. In striking contrast, the mechanisms driving their regulation have been less studied. Proteins involved in the GPI biosynthesis will be described next.

1.2.2 Biosynthesis of the GPI-anchor moiety

The biosynthesis of the GPI-anchor moiety occurs at least in eleven sequential steps in the endoplasmic reticulum (ER) by a series of more than 20 enzymes located in its membrane (Fig.3)

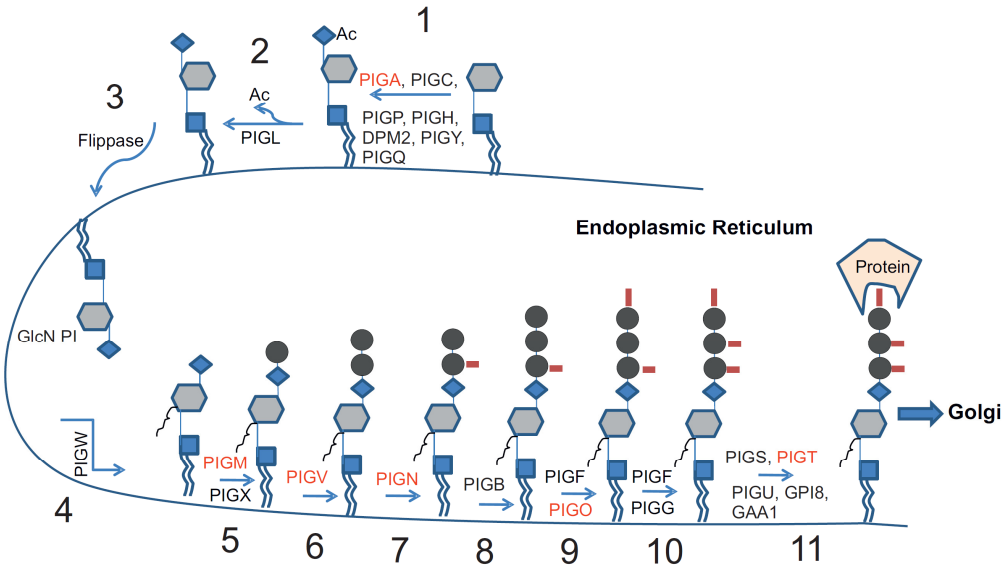


Fig. 4 – Biosynthetic pathway of the Glycosylphosphatidylinositol (GPI) anchor. Assembly of the GPI occurs in the Endoplasmic Reticulum and once assembled the anchor is attached to a protein containing the appropriate GPI attachment signal. The complex GPI-anchor protein (GPI-AP) is then exported to the cell surface via Golgi. Mutations in genes highlighted in red cause Inherited GPI Deficiency.

The steps in the biosynthesis of GPI are as follow:

Step 1: The synthesis of GPI is initiated by the addition of N-acetylglucosamine (GlcNAc) from UDP-GlcNAc to free PI on the cytosolic face of the ER. This reaction is mediated by GPI-GlcNAc, a large enzymatic complex constituted by PIGA (58-60), PIGC (61), PIGH (62), PIGQ (63), PIGY (64), PIGP (65) and two other additional components, PIGP and DPM2 (65). PIGA is the catalytic subunit of the complex.

Steps 2: De-acetylation of GlcNAc-PI is catalysed by PIGL, resulting in GlcN-PI (66).

Step 3: GlcN-PI is flipped to the luminal side of the ER by mediation of a yet unidentified enzyme, generally called "flippase".

Step 4: An acyl chain is transferred from an acyl-CoA to the 2-position of the inositol by PIGW, originating GlcN-(acyl)PI. The acylation of the inositol is indispensable for the attachment of the Etn-P to the terminal mannose later in the pathway (step 9) (67).

Steps 5, 6 and 8 – Constitute the addition of mannose residues. Three mannose residues from dolichol-phosphate-mannose (DPM) are subsequently added to GlcN-(acyl)PI through the action of specific GPI mannosyltransferases (GPI-MT), namely PIGM/PIGX (68, 69), PIGV and PIGB. GPI-MTI consists of two subunits: the catalytic subunit, PIGM, and the subunit required for its stabilisation, PIGX. The second and the third mannose residues are transferred by the GPI-MTII and the GPI-MTIII, PIGV and PIGB, respectively. A fourth mannose residue can also be added to the GPI intermediate by PIGZ (also called SMP3), depending on the mammalian cell type and tissue (68, 70-72).

Steps 7, 9 and 10 – Comprise Etn-P modification of mannose residues. The three mannose residues are modified by the addition of three Etn-Ps from phosphatidylethanolamine. The enzymes responsible for these modifications are the GPI-ethanolamine (GPI-ET) phosphate transferases. PIGF constitutes the stabilising element whereas PIGN, PIGO and PIGG have catalytic activity (73). Only the modification in the terminal mannose (Man-3) is essential for GPI expression in mammals (74).

Step 11 – Transamidation reaction. In the last step, a GPI transamidase complex composed by PIGK, PIGS, PIGT, PIGU and GAA1 cleaves the GPI attachment signal in the C-terminus of the precursor protein. The precursor protein in then bridged with the GPI anchor by the Etn-P attached to the terminal mannose (75-77).

Once assembled, GPI-APs are transported to the Golgi via secretory vesicles. Whereas in some circumstances of GPI deficiency, precursor proteins are retained in the ER and undergo ER-associated degradation, in others, even if they form an intermediate complex with GPI transamidase, they are hydrolysed and secreted in its soluble form. One such example is the secretion of free alkaline phosphatase (ALP) seen in patients with hyperphosphatasia mental retardation syndrome (HPMR) (78).

Recently, new sequencing technologies such as whole-exome sequencing analysis have allowed high throughput screening of potential disease-associated genes (79) and as a result, several novel diseases associated with GPI deficiency have been identified and characterised.

GPI- associated diseases comprise acquired- and inherited- GPI deficiency. Whereas acquired deficiency results from mutations in somatic cells, inherited deficiency is caused by germline mutations. By definition, a mutation is a pathogenic alteration in the DNA sequence typically with a population frequency of < 1% (80).

1.2.3 Acquired-GPI deficiency

Paroxysmal nocturnal haemoglobinuria (PNH), clinically defined in 1882, was the first identified GPI-associated disease (81). PNH is an acquired clonal disorder that usually arises from somatic mutations in the X-linked *PIGA* gene in one or more haematopoietic stem cells (HSCs) (60, 82), and therefore affects all blood cell lineages. So far, more than 100 somatic mutations in PIGA have been identified (83). Consistent with a clonal disorder, GPI deficiency is partial and presents in a multimodal fashion. Accordingly, three types of blood cells can co-exist: cells with normal GPI expression and cells with partial or complete GPI-deficiency (84). The expansion of GPI- clones in PNH suggests that other abnormalities in addition to *PIGA* mutations are involved in the pathogenesis of the disease and understanding how GPI- clones expand has been a subject of research in the last decades.

In PNH, the most dramatic effect is seen in the erythrocytes, where deficiency in the GPI-anchored complement inhibitors CD55 and CD59 triggers an unopposed activation of the complement cascade, leading to red cell destruction manifesting as intravascular haemolysis. The other defining features of PNH are venous thrombosis and bone marrow failure (BMF). The reason for bone marrow failure still remains unknown but various interesting findings have been made. The close association between PNH and aplastic anaemia (AA), in which stem cells in the bone marrow are destroyed by the autoimmune system has suggested that both diseases share a similar mechanism. Based on this hypothesis, selective targeting of GPI+ HSCs by autoreactive T –cells would permit GPI- HSCs to escape the immunological attack and to expand. In line with this, the analysis of the T-cell repertoire of PNH patients showed a high usage of the T-cell antigen receptor BV when compared to normal, age-matched controls (85) and more recently, Gargiulo, L. and collaborators, have shown the presence in PNH patients at frequencies higher than controls of a novel CD1d-dependent, GPI-reactive T-cell population, enriched in an invariant TCRα chain (86).
These findings support the hypothesis that GPI itself is a target of autoreactive T-cells in PNH and help to explain the pathogenetic mechanism driving BMF in PNH.

Defective GPI biosynthesis also constitutes the hallmark of inherited GPI deficiency (IGD), a group of rare autosomal recessive diseases caused by germline mutations in *PIGM*, *PIGA, PIGL*, *PIGV*, *PIGN*, *PIGO* and *PIGT* (Fig. 4).

1.2.4 Inherited-GPI deficiency (IGD)

Identified in 2006, *PIGM*-associated IGD was the first characterised inherited disorder of the GPI pathway (87). The disease results from a mutation in the core promoter of the first mannosyltransferase encoding-gene *PIGM* and so far, constitutes the only GPI-associated disorder resultant from an alteration in a regulatory region. Considering that my project relies on *PIGM*-associated IGD as a disease-model, I will describe it in more detail.

1.2.4.1 *PIGM*-associated IGD

PIGM-associated IGD was identified in two unrelated consanguineous families of Middle Eastern and Turkish origin (see Fig. 9), in which the probands presented in infancy with abdominal vein thrombosis and later developed absence seizures. Haematological investigations showed a normal thrombophilia screen in all affected children apart from an intermittently low-positive Ham test, indicating some degree of CD59 absence from the erythrocyte surface. Intriguingly, none of the affected children exhibited clinical evidence of significant intravascular haemolysis or bone marrow failure, defining features of PNH. Moreover, an unusual expression pattern of GPI-linked proteins and GPI itself on blood cells was observed. Whereas in PNH, distinct populations of cells resulting from normal, complete or intermediate loss of GPI are observed, in *PIGM*-associated IGD the expression pattern of GPI-anchored proteins varied according to the cell type. Red blood cells of the affected children presented near normal expression of CD59, with a small proportion of cells (<5%) being variably deficient while 30-50% of platelets were CD59 negative. Granulocytes were mostly GPI negative as assessed by fluorescein-labeled proaerolysin (FLAER), CD59 and CD24 antibodies. The constitutional nature of the disease was demonstrated by staining primary fibroblasts from one affected child with CD59 antibody and FLAER, in which there was a clear reduction in the expression of GPI.

By employing homozygosity mapping and genome-wide linkage search using SNP in both families, *PIGM* was identified as the only common candidate gene. As described above, PIGM along with PIGX, constitute the enzymatic complex GPI-MTI which is responsible for transferring the first mannose to GPI on the luminal side of the ER (68).

The block in GPI biosynthesis at the biochemical level was directly demonstrated by substrate labelling of patient-derived EBV-transformed lymphoblastoid B cell lines (LBCLs), followed by glycolipid extraction and high performance thin layer chromatography. Whereas the first steps of the biosynthetic pathway were intact, incorporation of H-D-mannose in GPI-enriched glycolipids was impaired in both families.

Altogether, those findings indicated that the defect arose at the addition of the first mannose moiety, pointing *PIGM* as the likely disease gene. This was further

confirmed in rescue experiments in which biosynthesis and expression of GPI at the cell surface were restored after transfection of PIGM-deficient LBCLs with PIGM WT cDNA.

Genetic defect in *PIGM*-associated IGD and its impact on transcription

In order to investigate the origin of the defect at the sequence level, Sanger sequencing analysis covering the single-exon of *PIGM*, its 3'- untranslated region (UTR) and 2kb region upstream of the ATG was performed. Sequencing analysis identified a single nucleotide substitution (C>G) at position -270 from ATG which was the cause of drastic reduction of *PIGM* mRNA levels in PIGM-deficient LBCLs (1% of normal controls). In heterozygous parental LBCLs, *PIGM* mRNA expression was about half of the normal levels which was consistent with a typical recessive trait.

TF binding motif prediction algorithms anticipated the presence of several putative transcription factor (TF) binding sites at the promoter of *PIGM* and very importantly, they predicted a putative Sp1-binding site encompassing the pathogenic -270C>G mutation (Fig. 5). Disruption of the Sp1-binding site by this mutation was further confirmed in EMSA and super-shift assays.

Fig. 5 - Predicted Sp1 binding sites at the *PIGM* promoter. Apart from the -270GC box, four putative Sp1binding sites were predicted at the promoter. -270C>G mutation (red arrowhead) abrogates Sp1 binding. Black arrow indicates the putative TSS.

The impact of the Sp1-binding mutation on promoter activity of *PIGM* was directly assessed in reporter assays. By transfecting PIG-M-deficient and heterozygous LBCLs with plasmids containing either the WT or the MUT promoter upstream of a luciferase reporter gene, a decrease of 80 to 90% in *PIGM* promoter activity was observed in the presence of the mutation (87).
Altogether, these findings identified the biochemical and genetic defect in *PIGM*-associated IGD and very importantly, they showed the importance of precise transcription factor binding and regulation of gene expression.

Sp1 belongs to the Sp/Kruppel family of transcription factors which are critical regulators of gene expression of many housekeeping and inducible genes (88-90). Mutations generating or abolishing Sp1 binding sites in regulatory regions are not novel and they have been reported in several human diseases (45, 91-94).
Further the molecular characterisation of *PIGM*-associated IGD, a therapy to treat patients was investigated.

From molecular characterisation to targeted therapy

Apart from interacting with the basal transcriptional machinery (95, 96) , Sp1is also known to influence transcription by recruiting and stimulating histone modifying enzymes, such as HATs (97, 98) and HDACs (99-101) to the gene promoter regions. The regulatory role of Sp1/Sp3 associated with the effect of the HDACs in promoting histone deacetylation and preventing high-order organization of the chromatin and the importance of the acetylation status during development prompted the authors to investigate whether *PIGM* promoter would contain responsive elements to butyrate and if so, how transcription of *PIGM* would be modulated.

By chromatin immunoprecipitation (ChIP) it was shown that in the presence of the -270 C>G mutation, histone 4 (H4) was hypoacetylated at the promoter in *PIGM*-deficient LBCLs and acetylated in heterozygous LBCLs. Treatment with 3mM sodium butyrate (NaBu), a pan- HDAC inhibitor, resulted in restoration of H4 acetylation at the MUT promoter, a 20-fold increase of the *PIGM* mRNA levels and restoration of GPI expression at the cell surface, as measured by FLAER. More importantly, in a child treated with sodium phenylbutyrate, peripheral blood cell *PIGM* mRNA levels drastically increased (from 0.15% before treatment to 60.77% after treatment), cell-surface GPI expression in granulocytes was restored to nearly normal levels and strikingly the child became seizure-free (102).

Altogether these results demonstrated that the -270C>G mutation disrupts a Sp1-dependent butyrate responsive element which is associated with histone hypoacetylation at the *PIGM* promoter. Further investigations regarding the molecular mechanism of transcriptional regulation in *PIGM*-associated IGD will be summarised in the next section.

As mentioned above, germline mutations in genes of the GPI pathway, other than *PIGM*, cause GPI deficiency. Remarkably, the main features shared between patients with IGD are the neurological abnormalities, including mental retardation and epilepsy. These features, frequently associated with an incomplete diagnosis, reinforce the significance of GPI and GPI-APs in normal neurodevelopment and

strengthen the importance of genetic screening and biochemical analysis of genes associated with glycosylation and in particular with the biosynthesis of GPI.

Interestingly, in this group of rare inherited disorders the severity of the disease not only correlates with the type of mutation but also with the location of the gene in the biosynthetic pathway. Accordingly, mutations in *PIGO* are more severe than mutations in *PIGV* since they cause a more pronounced growth delay in patients (103). Other very interesting evidence in this group of disorders is the absence of haemolysis in patients, suggesting normal GPI expression in the surface of the erythrocytes. Understanding how those genetic abnormalities contribute to different phenotypic patterns in the haematopoietic cells still remains to be elucidated. The contribution of somatic events in this group of rare diseases should be also start taken into account.

The genetic abnormalities identified so far in genes of the GPI biosynthetic pathway as well as the associated clinical features are summarized below.

Table 2 – Genetic abnormalities identified in the GPI biosynthetic genes and associated clinical features

Gene	Mutation	Disease	Clinical features	References
PIGM	Homozygous g.-270G>C	*PIGM*-associated IGD	Epilepsy, thrombosis of the hepatic veins	(87) (102)
PIGA	c.1234C>T	X-linked disease	Cleft palate, neonatal seizures, contractures and central nervous system (CNS) malformations	(104)
PIGL	Compound heterozygous c.500T>C + c.274delC, c.652C>T, c.427-1G>A or del17p12-p11.2	CHIME	Coloboma, heart defects, ichthyosiform dermatosis, mental retardation and ear anomalies with hearing loss	(105) (106) (107) (108) (109)
PIGV	Homozygous c.1022C>A or c.766C>A; compound heterozygous c.1022C>A+c.1154A>C; c.1022C>A+c.1022C>T	HPMR	Mental retardation and elevated alkaline phosphatase (ALP) levels	(78) (110) (111, 112)
PIGN	c.2126G>A	MCAHS (multiple-congenital-anomalies-hypotonia-seizures syndrome)	Developmental delay, cardiac, genitourinary and gastrointestinal anomalies, severe neurological impairment with chorea and seizures	(113)
PIGO	Compound heterozygous c.2869C>T+ c.2361dup; c.2869C>T+ c.3069+5G>A	-	Mental retardation, elevated alkaline phosphatase *TNAP* gene, intractable seizures Growth delay, malformations of the urinary system and heart	(103)
PIGT	Homozygous c.547A>C		Distinct facial features, intellectual disability, hypotonia and seizures, abnormal skeletal, endocrine and ophthalmologic findings	(114)
	Compound heterozygous:c.1401-2A>G; 8-MB deletion	PNH	Hemolytic crisis, abdominal pain, diarrhea, headache, arthralgia, dyspnea and fatigue	(115)

1.2.4.2 Deregulation of *PIGM* transcription

Gene silencing of CGI promoters is achieved through dense CpG methylation as mentioned before, or Polycomb (PcG) recruitment. Whether CpG methylation is the primary event in gene silencing is still a matter of debate. The PcG complex in mammals comprises polycomb-repressive complex 1 (PRC1) and 2 (PRC2). PRC2 is recruited to CGIs and then trymethylates H3K27 which is recognised by PRC1 complexes that impede transcriptional elongation, thus repressing gene expression by further modifying the histone tails (H2A-K119 ubiquitination) (116-118). The hierarchy of functional regulators that is responsible for Polycomb recruitment is still under investigation. Most results have arisen from studies in embryonic stem (ES) cells in which gene expression activation and repression has to be properly controlled.

Taking advantage of the -270C>G mutation at the CGI-*PIGM* promoter, the transcriptional mechanism driving repression of the housekeeping gene was investigated in more detail.

Having seen that in B cells Sp1 binding was required for *PIGM* transcription and histone acetylation, the role of the CGI in the nucleosome dynamics was further evaluated. CGIs are intrinsically associated with nucleosome destabilization, accessible and nucleosome-free chromatin and therefore with transcriptional activation of gene expression (116, 117). The binding motifs of several TFs, including of Sp1, Nfr-1, E2F and Ets, are GC-rich and overlap with GCIs (119).
Restriction enzyme accessibility assay (REAA) showed that the MUT promoter was less accessible than the WT promoter, consistent with nucleosome compaction. Nucleosome positioning, determined by micrococcal nuclease (MNase) protection assay followed by RQ-PCR showed low levels of protection in the WT promoter, particularly in the area encompassing the TSS and the -270GC box thus defining the nucleosome-depleted region (NDR). In the MUT promoter, high levels of protection seen across the NDR were reduced by restoration of Sp1 binding following NaBu treatment, and most likely by histone re-acetylation. Together, the results suggested that the CGI was not sufficient to impose a nucleosome depleted region (NDR) at the *PIGM* promoter and intact

Sp1 binding and recruitment of HATs were required for nucleosome relaxation and normal transcription.

Repression of gene expression by nucleosome compaction is usually associated with DNA and histone methylation. Accordingly, the MUT promoter was enriched for the H3K27me3 mark, characteristic of developmentally regulated/poised genes, and suggestive of PcG recruitment and transcriptional repression. Further analysis showed that the CGI promoter was not methylated. Surprisingly, the co- presence of the active mark, H3K4me3 in the MUT promoter suggested the presence of a bivalent chromatin state. Altogether, the findings indicated that active TF binding "protects" against silencing by PcG, instructing a permissive chromatin state. Contrary, absence of active TF binding in CGIs directs Polycomb recruitment and deposition of bivalent histone methylation marks thus explaining the repression, yet reversible, of *PIGM* transcription and GPI deficiency in B cells (38).

Understanding the mechanism by which erythroid cells overcome the pathogenic C>G mutation and express near normal levels of GPI and GPI-APs demanded further investigation.

Below, as a prelude to the experimental work I plan to address this aspect of the molecular pathogenesis of *PIGM*-associated IGD, I provide a brief overview of the erythropoiesis and its transcriptional control. Before, a summary of the role of Sp TFs in transcriptional regulation is presented.

1.3 The Specificity Protein (Sp) transcription factor family

The Specificity Protein/Krüppel-like Factor (Sp/KLF) transcription factor family shares a highly conserved DNA binding domain of C_2H_2-type zinc fingers. Members of this family bind to CACCC/GC/GT boxes, which are very common in the regulatory regions of many cellular genes.

The Sp subgroup favours GC-rich boxes and so far, nine members have been identified, Sp1 to Sp9. The glutamine-rich subfamily is constituted by Sp1, Sp2, Sp3 and Sp4. Whereas Sp1 and Sp3 are ubiquitously expressed, Sp4 is cell-type specific and presents distinct functional properties from Sp1 and Sp3. Sp4 is highly enriched in the brain tissue (120, 121) and therefore, Sp4 deficiency results

in abnormalities more prominent in the nervous system (122). In fact, alterations in the Sp4 and Sp1 levels have been associated to neuropsychiatric disorders (123) (124) and a very recent study proposes these proteins as potential biomarkers of early-stages of psychotic diseases (125).

Depending on the promoter architecture and cellular context, Sp1 and Sp3 compete for DNA binding and it has also been suggested that Sp1 and Sp3 are functional redundant. However, results derived from developmental studies in mice are not in agreement and suggest they might have functional redundant roles only in early stages of development. Accordingly, Sp1 knockout mice are not viable and die around embryonic day 11 (126) whereas Sp3 knockout mice are viable and die from respiratory abnormalities (127).

Sp2 contains the least conserved DNA binding domain within the Sp family and mainly recognises GC box promoter elements. A recent study has identified various Sp2 gene targets in mouse embryonic fibroblasts (MEFs) and human embryonic kidney (HEK) 293 cells, defining Sp2 as a major regulator of genes involved in various cellular pathways (128).

Sp1 was the first identified member of the Sp family and many of its functions have been revealed by *in vitro* experiments. Sp1 has two isoforms and it is constituted by four domains. Two transactivation domains, that are localised in the N-terminus and are required for Sp1 transactivation potential; an internal domain and a carboxyl-terminus domain required for synergistic transactivation (Fig. 6). The N-terminus domains directly contact elements of the transcriptional machinery, such as TBP (TATA-binding protein)/TFIID (95, 96), thus activating transcription initiation. It is possible that in this way Sp1 directly recruits TFIID that potentiates Sp1-mediated transcription of TATA-less genes. Interestingly, Sp1 also has the ability to regulate gene expression by mediating chromatin loops between enhancer and promoter regions (129).

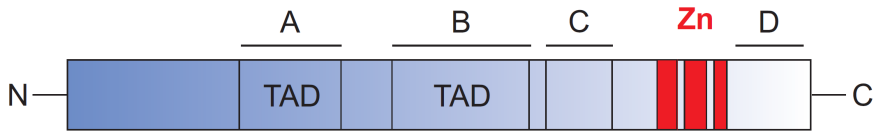


Fig. 6 – Schematic representation of Sp1. Sp1 is constituted by four domains (A, B, C and D). Domains A and B comprise the transactivation domains (TADs). Sp1 binds to the DNA through their zinc-fingers (Zn).

Sp1 is a very versatile protein, regulating expression of genes involved in various cellular processes, such as cell differentiation (130), apoptosis (131), angiogenesis and cancer (132). The posttranslational modifications of Sp1 affect its transcriptional activity and stability and therefore contribute for the complexity of its regulatory functions (133). Among those modifications are phosphorylation, acetylation, sumoylation, ubiquitylation and glycosylation.

For a long time, Sp1 was recognised as an activator of transcription of housekeeping and TATA-less genes. However, it is now known that Sp1 can also activate transcription of tissue-specific genes by interacting with tissue specific TFs such as GATA-1 (134, 135). Interactions with signal-induced factors, like SREBP-1a have also been reported (136).

As mentioned before, Sp1 interacts with HATs and HDACs.

Mutations in DNA sequences causing disruption of Sp binding sites have been reported in a growing body of literature. Notably, these mutations have an impact in normal gene expression and are in the basis of various human diseases.

In 1991, Toshiyuki *et al.* (137) described for the first time a Sp1-binding site mutation in the retinoblastoma gene. Koivisto *et al.* showed in 1994 that a mutation (-43C>T) in the core promoter of the low density lipoprotein (LDL) receptor gene abolishes binding of Sp1 and causes familial hypercholesterolemia (138). Posteriorly, another study identified a mutation in a Sp1 binding site in the 5'-upstream region of the *DKC1* gene, causing multisystem disorder dyskeratosis congenita (DKC) (91), associated later on with decreased promoter activity and thus, gene expression (92). More recently, it has been shown that a mutation (-250G>C) in the promoter of the ferrochelatase gene (*FECH*) abrogates binding of

Sp1 and impairs promoter activity causing erythropoietic protoporphyria (EPP) (93).

Abrogation of Sp1 binding in the regulatory region of the *NME4* gene impairs binding of the erythroid specific TF SCL/Tal1 localised 55bp apart. Consequently, *NME4* expression is altered in erythroid cells but is not affected in nonerythroid cells (139).

These results constitute a remarkable evidence of how alterations in gene regulatory regions have a functional impact in TF regulatory networks of the cell, affecting gene expression in a more broad or cell-type specific manner.

1.4 Haematopoiesis

Haematopoiesis is the biological process in which a pluripotent haematopoietic stem cell (HSC) gives rise to haematopoietic progenitor cells which in turn differentiate in different blood cell types. In mammals, haematopoiesis occurs sequentially in the yolk sack, fetal liver and bone marrow.

A pluripotent HSC is highly proliferative and capable of self-renewal, or in other words, to produce new HSCs. After differentiation, a long-term HSC ((LT)-HSC) originates an HSC with lower self-renewal potential, designated short term (ST)-HSC.

(ST)-HSC has the ability to differentiate into a multipotent progenitor cell–MPP, which in turn differentiates into a common myeloid progenitor (CMP) or common lymphoid progenitor (CLP) cell. CMP then differentiates into megakaryocyte/erythroid progenitor, called MEP or into granulocyte/macrophage progenitor, GMP; CLP originates B and T lymphocytes. MEPs give rise to red blood cells (RBC) and megakaryocytes and GMPs give rise to mast cells, eosinophils, neutrophils and monocytes/macrophages (Fig. 7). The specific pathway that gives rise to red blood cells is called erythropoiesis and it will be described next. Specific markers at the cell surface of the different haematopoietic cell types allow their identification and selection. For instance, CD34, c-kit, IL-6R, Thy-1 and CD45RA constitute cell surface markers of HSCs (140).

External signals such as cytokines (SCF, TPO, Flt3 and GM-CSF) produced by stromal bone marrow cells, are known to play a critical role in the status of HSC and therefore influence haematopoiesis.

During the haematopoietic development, TFs play a crucial role by influencing the transcriptional programmes in both the formation of the HSC and maintenance of its function and in the blood cell lineage specification. In addition to binding to DNA, haematopoietic TFs interact with each other and also interact with chromatin-associated factors.

Haematopoietic TFs have to be expressed in an ordered specific manner. Amongst the transcription factors essential for the survival and proliferation of HSCs are

MLL (mixed lineage-leukemia gene), RunX1, TEL/ETV6, SCL/Tal-1 and LMO2 (141). SCL/Tal-1 and LMO2 are essential in the early stages of haematopoiesis as mice null embryos die from the complete absence of embryonic blood (142).

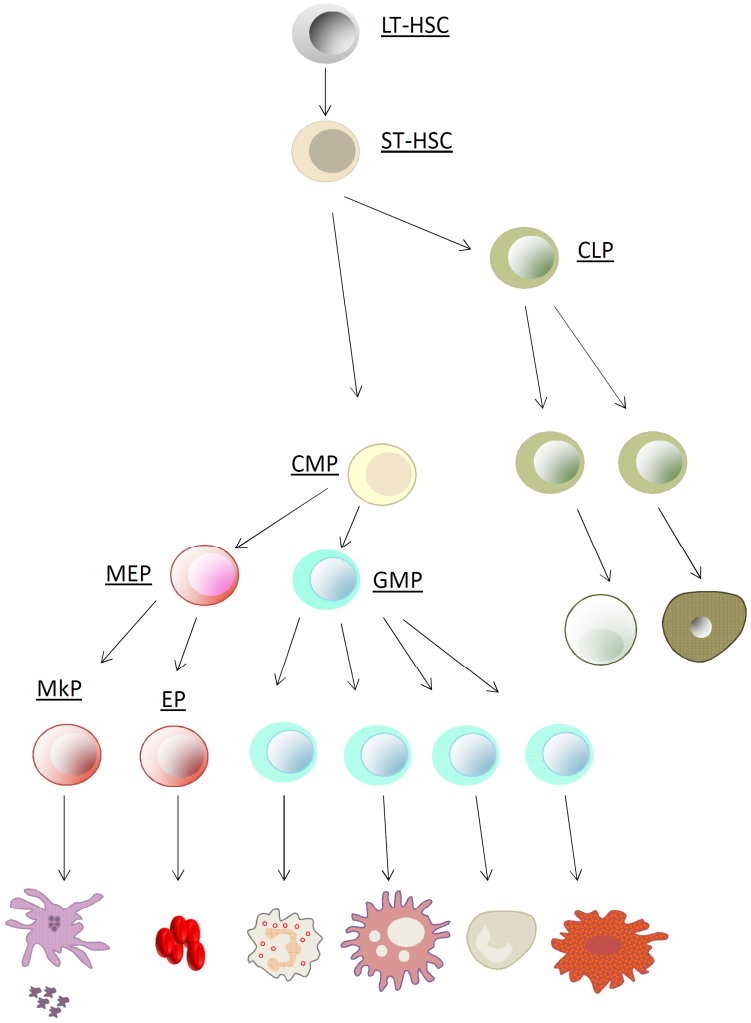


Fig. 7 – Haematopoietic hierarchy. The long-term HSC (LT-HSC) origins a short-term (ST-HSC) which in turn gives rise to common-myeloid progenitor (CMP) and common-lymphoid progenitor (CLP). CMP can origin the megakaryocyte/erythrocyte progenitor (MEP) and granulocyte/macrophage progenitor (GMP) precursor cells. Whereas MEP origins megakaryocytes and erythroid precursors, GMP gives rise to granulocytes, mast cells, eosinophils and monocytes/macrophages. CLP differentiates into B and T lymphocytes (adapted from(141)).

In what respects lineage specification, TFs play a dual role. They promote lineage differentiation and at the same time they act against factors favouring other choices. Accordingly, PU.1, a key TF for the monocytic and B lymphoid commitment, interacts and antagonizes GATA-1, which is essential in the megakaryocytic and erythroid development (143); the contrary is also true (144).
Direct evidence of the importance of TFs in haematopoiesis has arisen from genetic studies using knock-out mice. One such example is the master regulator of erythropoiesis, GATA-1. Disruption of GATA-1 in a murine embryonic stem cell line failed to give rise to red blood cells in chimaeric mice (145) and in its absence, erythroid precursor cells arrest at the proerythroblast stage and undergo apoptosis (146).
Forced expression studies and reprogramming, also revealed the importance of master TFs in lineage specification. For instance, exogenous expression of GATA-1 in cell lines of monocytes induced the expression of erythroid-megakaryocyte lineage markers and down-regulation of monocytic markers (147, 148). Overexpression of C/EBP-α, was also shown to commit B- and T-cell progenitors into functional macrophages (149, 150).

Not surprisingly, de-regulation of TFs and chromatin-modifying enzymes in the haematopoietic cells is largely associated with the onset of a variety of haematological malignancies. For instance, heterozygous mutations in PU.1 have been linked to acute myeloid leukemia (151) and overexpression of Tal-1, and its consequent abnormal auto-regulation cause T- cell acute lymphoblastic leukemia (T-ALL) (152).

1.4.1 Erythropoiesis

Erythropoiesis is the process in haematopoietic cell lineage specification that gives rise to erythrocytes. As the erythroid precursor cells undergo differentiation their numbers increase and their proliferative potential decreases. During fetal life, erythropoiesis occurs in the fetal liver and following birth, switches to the bone

marrow. In adult life, erythroid precursors are mainly produced in the bone marrow and once they mature, they expel their nucleus and enter the bloodstream. The growth of burst forming unit-erythroid (BFU-E), the more immature erythroid progenitors, depends of cytokines such as stem cell factor (SCF), thrombopoietin (TPO), interleukin 3 (IL3) and 11 (IL11), and Flt-3 ligand. Under normal conditions these cells differentiate into colony-forming unit-erythroid (CFU-E) cells which are highly dependent of erythropoietin (EPO) and give rise to the erythroid precursors. At this stage, pro-erythroblasts undergo four to five cell divisions and sequentially originate morphologically distinct populations - basophilic, polychromatic and orthochromatic erythroblasts -, through acquisition of the surface markers, CD36, transferrin receptor protein 1 (CD71) and glycophorin A (GPA). CD71 loss anticipates enucleation, giving rise to a more mature and reduced in size cell, the reticulocyte (Fig. 8). Reticulocytes then enter the blood circulation where they mature into erythrocytes. The life-span of erythrocytes is 120 days (153). Erythroid gene regulation not only depends of the antagonising effect between haematopoietic TFs, but also depends of chromatin changes mediated by histone post-translational modifications (154).

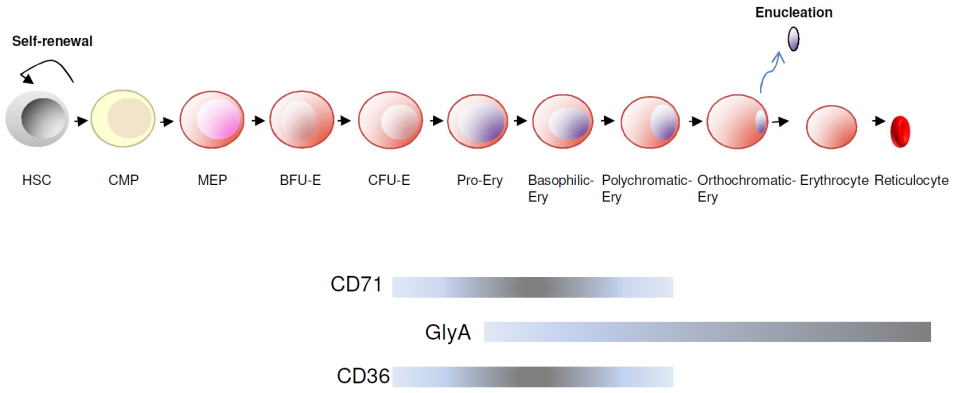


Fig. 8 – Erythroid differentiation. HSC – Haematopoietic Stem Cell; CMP – Common Myeloid Progenitor; MEP – Megakaryocyte-Erythroid Progenitor; Ery – Erythroblast. Erythroid cells are identified at the various stages by expression of CD71, GlyA and CD36. Adapted from (153).

1.4.2 Regulation of erythropoiesis by GATA-1 and KLF1 Transcription Factors

During haematopoietic differentiation, lineage specification is largely determined by the coordinated expression of specific TFs, in a quantitative and temporal manner.

The GATA family of TFs has been implicated in the specification of erythroid/megakaryocytic cells as well as in the transcriptional program of T cells (155). The GATA family comprises the haematopoietic and non-haematopoietic subfamily. The haematopoietic sub-family is constituted by GATA-1, GATA-2 and GATA-3 and their expression is restricted to the haematopoietic system (156); the non-haematopoietic sub-family comprises GATA-4, GATA-5 and GATA-6 which are expressed in several tissues (157).

GATA-1 constitutes the master regulator of erythropoiesis as it is responsible for survival and proliferation of erythroid cells. It is also expressed in megakaryocytes, eosinophils and mast cells. GATA-1 has been extensively studied both in mouse and human erythroid cells. GATA-1 deficient embryonic stem cells arrest at the proerythroblast stage and GATA-1 null mice show complete ablation of erythropoiesis and die from severe anaemia (158).

GATA-1 is constituted by two zinc-fingers: a C-terminal zinc-finger that binds DNA with high-affinity and an N-terminal zinc finger responsible for stabilizing the interaction (159, 160). GATA-1 regulates transcription of erythroid-specific genes through binding to its consensus motif (T/A)GATA(A/G) in gene regulatory regions. Recently, ChIP-seq data in a mouse-erythroid cell line identified more than 2,000 GATA-1 targets spanning more than 15,000 binding sites (161). It usually regulates gene expression in association with FOG-1 (162), Scl/Tal-1 (163), LMO2 (164) and others. It also interacts with TFs like KLF1 (165) and proteins with acetyltransferase properties like CBP/p300 (166, 167).

GATA-1 and Sp1 binding sites are very often localised in close proximity in the promoters and enhancer of erythroid-specific genes. GATA-1 physically interacts

with Sp1 through their zinc finger domains and in the absence of GATA-1 binding sites, Sp1 is able to recruit GATA-1 at the promoter (134, 135, 168).
Studies performed in vitro revealed that depending on the promoter context, GATA-1 functional interacts with Sp1 or with its close associated family member, KLF1. Accordingly, GATA-1 cooperates with Sp1 at the pyruvate kinase promoter and with KLF1 at the β^{maj} globin-derived GCT promoter. At the glycophorin B promoter, GATA-1 interacts with either Sp1 or KLF1 (165).

Disease associated mutations in GATA-1 have also been described, revealing the multifactorial role of GATA-1. A mutation in a splice site of the GATA-1 gene produces a truncated short-length form of the protein causing Diamond-Blackfan-Anemia (DBA) (169). Similarly, GATA-1 mutations that originate an amino-terminally truncated protein contribute for the onset of transient myeloproliferative disorder (TAM) and acute megakaryoblastic leukaemia (AMKL) in Down-Syndrome patients (170, 171).

The KLF subgroup of the Sp/KLF transcription factor family is comprised by 16 members (KLF1-KLF16) and preferentially binds to CACCC boxes. KLF1, also known as EKLF, has restricted erythroid expression and therefore regulates erythroid genes such as the β-globin (*HBB*) gene. Mutations in the CACCC box in the regulatory region of β-globin are a cause of β-thalassemia (172). KLF1 is constituted by three zinc-fingers and mutations in those have been directly associated with hereditary persistence of fetal haemoglobin (HPFH) (173) and congenital dyserythropoietic anemia (CDA) (174). In mice, KLF1 knockout cause defects in the late stages of fetal erythropoiesis (175, 176). The second member of the KLF family, KLF2 is specifically expressed in T lymphocytes.
Similarly to Sp1 and Sp3 also KLF1 and KLF3 (BKLF) compete for the same sites. Interestingly, Sp1 has the ability to compete with KLF family members such as KLF4, KLF9 and KLF13 for the same binding sites (177).

1.5. Aims and Hypothesis

Mutations in gene regulatory regions are less frequent than coding mutations and therefore less well studied. However, when they occur, they provide unique opportunities for elucidating mechanisms of transcriptional regulation and their functional impact in gene expression.

PIGM-associated IGD provides the opportunity to elucidate the role of a *cis*-DNA element, in this case of the -270GC-rich box, in differential, tissue-specific transcriptional regulation of the same gene. As discussed earlier, in *PIGM*-associated IGD the -270 C>G promoter mutation differentially affects expression of GPI in cells from the same or different tissues, as exemplified by erythrocytes which are mostly GPI+ and by the granulocytes which are mostly negative for GPI expression.

The molecular events underlying differential regulation of *PIGM* and variable expression of GPI in haematopoietic cells are not known and their elucidation constitutes the overall aim of this thesis.

The **specific aims** of this thesis are to:

- Elucidate the differences at the *PIGM* transcriptional level between normal and IGD haematopoietic cells and to investigate the *PIGM* promoter architecture (nucleosome positioning and promoter accessibility) and its impact in gene expression;
- Determine the role of the mega-erythroid lineage-affiliated transcription factors GATA-1 and KLF1 in transcriptional regulation of *PIGM* in erythroid cells and
- Determine the role of the generic transcription factor Sp1 and other Sp family members on *PIGM* transcriptional regulation in normal and IGD haematopoietic cells;
- Explore the role of the pathogenic -270C>G mutation in the interactions established between *PIGM* and other genes.

Chapter 2 - Materials and Methods

2.1 Cell lines, primary cells and treatment

K562 (erythroleukemia), HL60 (myeloid), Jurkat (T-lymphocyte) and EBV-immortalised lymphoblastoid B cells (LBCLs), detailed in Table 2, were cultured in RPMI-1640 medium (Sigma). 293T, Flp-In-293, HeLa and SH-SY5Y cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM; Sigma) and splited by trypsinisation. Both media were supplemented with 10% fetal bovine serum (FBS; Sigma), 1% L-Glutamine (Sigma) and 1% penicillin-streptomycin (Sigma).

Table 3 - EBV-immortalised lymphoblastoid B cells (LBCLs) (see also family tree below)

Cell line	Genotype	Family member
N1	Homozygous WT	N/A
N2	Homozygous WT	N/A
1A	Homozygous WT	IV1, Family 1
1B	Homozygous MUT	IV4, Family 1
2E	Heterozygous	III4, Family 2
2A	Heterozygous	III5, Family 2
2B	Homozygous MUT	IV2, Family 2
2C	Homozygous MUT	IV3, Family 2
2D	Heterozygous	IV4, Family 2

N1, N2 – EBV transformed cell lines, Triose-Phosphate Isomerase (TPI) deficient.

Primary cells were obtained from healthy donors and from *PIGM*-associated IGD patients and their relatives, as reported in (87), from which this current study follows. PBMC from patients and relatives have been cryopreserved since then.

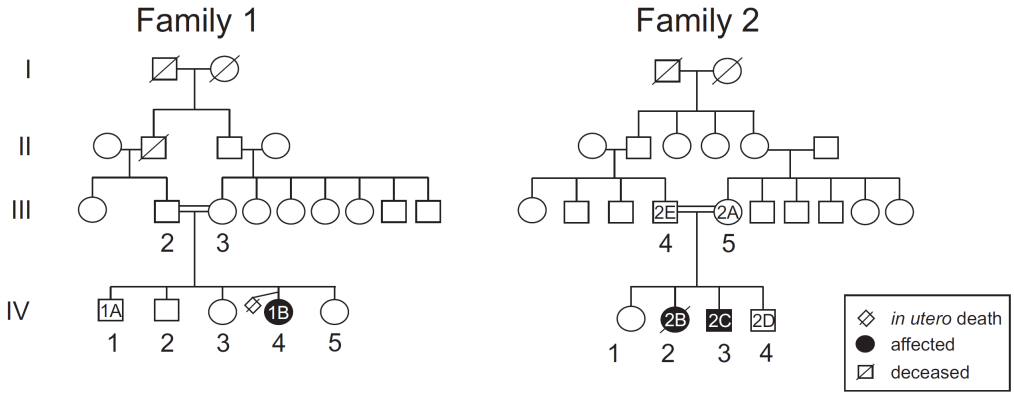


Fig. 9 – Family trees of children with inherited GPI deficiency. In Family 1 the parents are first cousins and there are four unaffected siblings and one affected child (1B); her twin sibling died *in utero* at 8 weeks. In Family 2 the parents are first cousins and there are four children, two of whom are affected (2B and 2C).

Peripheral blood mononuclear cells (PBMC) were isolated by density gradient centrifugation at 1800 rpm for 30 minutes using Histopaque (Sigma). $CD34^+$ cells were obtained from cord blood mononuclear cells, after density gradient centrifugation, by using the CD34 MicroBead cell isolation kit (Miltenyi Biotec). After isolation, $CD34^+$ cells were analysed for purity in a BD LSRFortessa flow cytometer (BD Biosciences, USA).

Neutrophils were obtained from peripheral blood by lysis of the erythrocyte fraction in sequential incubations of ice-cold ddH_2O for 10 seconds, 0.6M KCl and ice-cold PBS, spun at 1300 rpm, 4ºC, for 8 minutes. Cell lysis was performed until only a clear white pellet of neutrophils remained. Cell morphology was monitored by May-Grünwald-Giemsa stained cytospin preparations (as described in 2.4).

Mithramycin (MitA; Sigma) was dissolved in DMSO at 50µM stock solution and diluted prior to use at the indicated final concentrations. After 24 hours, viable

cells (as determined by trypan blue (Sigma) exclusion were counted in a hemocytometer).

2.2 Bioinformatics

Nucleotide sequences were obtained from the Ensembl database (http://www.ensembl.org/index.html) and analysis of the *PIGM* promoter region was carried out using the Transcription Element Search System (TESS) and the TFsearch (http://www.cbrc.jp/research/db/TFSEARCH.html) web tools. ChIP-seq data was retrieved from the ENCODE (http://genome.ucsc.edu/encode/) database. Primers were designed with NCBI Primer Blast.

2.3 Bacterial transformation

Chemically competent bacteria were thawed on ice and 50ηg of plasmid DNA was added followed by incubation on ice for 30 minutes. Transformation was carried out by heat-shock at 42ºC, for 45 seconds. 250µl of LB medium (Sigma) was added to the cells and they were shaken at 225 rpm, 37ºC for 1 hour, following which 50µl were plated in LB agar plates containing 100µg/ml ampicillin (Sigma) and incubated O/N at 37ºC. Individual colonies were picked, seeded into 4ml LB medium containing 100µg /ml ampicillin and shaken at 225 rpm, 37ºC, O/N. Plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit (Fermentas).

Larger quantities of plasmid DNA were obtained by seeding the bacterial cells into 300ml of LB medium containing 100ug/ml ampicillin, shaken at 225 rpm, 37ºC, O/N. Plasmid DNA was isolated using the GeneJET Plasmid Maxiprep Kit (Fermentas).

2.4 Cytospins preparations

Approximately $4x10^4$ cells were placed in a cytospin cartridge and spun at 400 rpm for 5 minutes into a slide (Superfrost Plus,VWR). Slides were air-dried, fixed in methanol (Sigma) and incubated in the May-Grünwald solution (VWR)

previously diluted 1:2 in H_2O, for 7 minutes. Slides were then incubated in the Giemsa solution (VWR) previously diluted 1:10 in H_2O, for 20 minutes and washed 3 times in distilled H_2O. Slides were mounted in DPX solution (Sigma) and analysed in a light microscope (EVOS XL Core, Fisher Scientific, USA).

2.5 Chromatin Immunoprecipitation (ChIP) assays

Cells lines ($5x10^6$/IP) and $CD36^+$ primary cells ($3x10^5$/IP) were collected by centrifugation at 1200 rpm and cross-linked with 1% formaldehyde (Sigma) at RT for 10 minutes. Cross-link was quenched by addition of glycine to a final concentration of 125mM. According to the antibody, chromatin was prepared using the Shearing ChIP kit (Diagenode) or using the solutions recommended in the ChIP assay kit (ref 17-295, Millipore) and sheared by sonication at 4ºC for 25 minutes in a 0.5/0.5 ON/OFF cycle, high intensity in a Bioruptor UCD-200 (Diagenode, Belgium). Sonicated fragments of average 500bp length were confirmed on a 1.5% agarose gel. Preblocked (0.2mg/ml glycogen, 0.2mg/ml BSA) Protein G magnetic beads (Dynabeads, Invitrogen) were then incubated with 2-5µg each antibody (Table 4) for at least 2 hours, at 4ºC in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2M EDTA, 16.7mM Tris-HCl pH8.1, 150mM NaCl) containing protease inhibitors (PI, Sigma). Chromatin was subsequently diluted at least 10 times in ChIP dilution buffer (plus PI) and pre-cleared by incubation with Protein G magnetic beads for 1 hour at 4ºC on a rotating wheel. Pre-cleared chromatin was incubated with the antibody-beads complex overnight at 4ºC on a rotating wheel. 1/10 of the amount of chromatin used per IP was kept as Input control. Protein-DNA complexes were then recovered and sequentially washed in ChIP dilution buffer, Low Salt Wash Buffer (50mM Hepes pH7.9, 500mM NaCl, 1mM EDTA, 1% Triton X-100), High Salt Wash Buffer (20mM Tris-HCl pH8, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 500mM NaCl) and TE (10mM Tris-HCl pH8, 1mM EDTA), at 4ºC for 5 minutes. Protein-DNA complexes were eluted by adding 150µl elution buffer (1% SDS, 0.1M $NaHCO_3$) at 65ºC for 4 hours and a second time for 30 minutes. Eluted samples and Input control were treated with 200µg/ml Proteinase K (Fermentas). DNA was recovered by phenol-chloroform extraction and ethanol precipitation.

RQ-PCR was performed using specific primers for the target regions (Table 4) and Maxima SYBR Green master mix (Fermentas) in an Applied Biosystems 7500 Real-time PCR system (Applied Biosystems, USA). The PCR program used for amplification consisted of a denaturation step at 95ºC for 10 minutes, followed by 42 cycles of 2 minutes at 94ºC, 30 seconds at 60ºC and 1 minute at 72ºC and melting curve. Enrichment of the target sequence was calculated as % of Input using the formula: [(2^ -Ct IP)/(2^ -Ct Input)]*(100/DF), where DF is the dilution factor of the Input. Data was presented as fold enrichment relative to % Input IgG.

Table 4. Antibodies used for ChIP

Antibodies	
GATA-1	ab11852, abcam
KLF1	ab2483, abcam
Sp1	17-601, Millipore
Sp2	sc-643, Santa Cruz
IgG rabbit	sc-2027, Santa Cruz
IgG mouse	sc-2025, Santa Cruz
IgG goat	sc-2028, Santa Cruz

Table 5. ChIP-RQ primers

Designation	Forward primer	Reverse primer
8	5'–GGATGGCAACTTTGACAGAAGAAA- 3'	5'–TTCCTTAACTTGTCAGCCCCTTA-3'
7	5'-TCACCTTCACCTAGGGTTGC-3'	5'-AGTTGAAGGGGGCTGAGTGT-3'
6	5'-CCACAAATCCCCGGTGAATAC-3'	5'-CTGCGCGCTACATCCTTCAG-3'
5	5'-TCACTCTCGGATCTCCAGC-3'	5'-GAGATTTCTTTGTAAGTGACAGCC-3'
4	5'-GCTGTCACTTACAAAGAAATCTCAC-3'	5'-CTCGCTTCCGCTTCTTCTTC-3'
3	5'-CAAAGAAATCTCACTCCACTCCA-3'	5'-CGAGCCAAAAACTTGCCTTC-3'
2	5'-GAAGAAGCGGAAGCGAGAC-3'	5'-CTATCACATCCGGCATGAAG-3'
1	5'-CTGGAGACAGGTAGCAGTACG -3'	5'--CTTCCCTTCGGTTCTTTGCAG-3'
+6Kb	5'-TCTTTTGAGGGGCAGTTCTCTC-3'	5'-TTTGGTGAGAGAGGGTTTGGAG-3'
HS-49	5'-TCTCCTTGTAGTTGCCACAAGCT-3'	5-GGGTGGGCTATAAGTGACTAAAAGAG-3'
HS-3.5	5'-TGATCTGGCCCCACTGATTC-3'	5'-AGAAGTCTGCGGCGACAGTT-3'
HS+14	5'-TTATCTCTCCGGTCCCAGCTT-3'	5'-GCCGGTTTGTCCTCCTAATCT-3'
HS+15	5'-GCCGGTTTGTCCTCCTAATCT-3'	5'-GACGCAGCCCCTTTCTCAA-3'
DHFR	5'-TCGCCTGCACAAATAGGGAC-3'	5'-AGAACGCGCGGTCAAGTTT-3'
HBA2	5'-GGGCCGGCACTCTTCTG-3'	5'-GGCCTTGACGTTGGTCTTGT-3'
GW10	5'-GGCTAATCCTCTATGGGAGTCTGTC-3'	5'-CCAGGTGCTCAAGGTCAACATC-3'
Sp2 -2.5Kb	5'-TACAGGTGAGTCATGGAACC -3'	5' –ACCCTGCTCAAAGTCTGCAT -3'
Sp2	5' –CCGTCCATCAGTGACATTGC -3'	5' –TACGACATCTTCCTCCCTGG -3'

2.6 Circular chromosome conformation capture (4C)

Cells ($1x10^7$/condition) were centrifuged for 5 min at 1500 rpm at RT and washed twice with PBS. Pellet was ressuspended in 500µl of 10% (v/v) FCS/PBS and then cross-linked in 1% formaldehyde (Sigma) for 10 minutes at RT. Cross-link was quenched on ice by addition of glycine to a final concentration of 0.125M. In parallel, an uncross-linked sample was prepared. Cells were centrifuged for 8 min at 230g, 4°C, and cell pellet was ressuspended in 5 ml of cold lysis buffer (10mM Tris-HCl, pH7.5; 10mM NaCl; 5mM $MgCl_2$; 0.1mM EGTA; 0.2% NP-40; 1xPI) and incubated for 10 min on ice. After assessing cell lysis by trypan blue, the lysate was centrifuged for 5 min, 400g at 4°C, and then ressuspended in 500µl of 1.2X restriction enzyme buffer. 7.5µl of SDS was added (0.3% final concentration) and nuclei were incubated at 37°C for 1 hour, 1400 rpm. 50µl of Triton-X 100 (2% final concentration) was added to block SDS for 1 hour at 37°C, 900 rpm. A 5µl aliquot was taken as undigested genomic DNA control. Digestion was carried out by adding Apo I enzyme (Fermentas) for the following period: 100U – 1 hour, 300U – O/N and 100U – 1 hour, at 37°C, 1400 rpm. An aliquot (15µl) was taken as control of digestion and verified on a 1.5% agarose

gel. Prior to ligation, samples were incubated with 40µl of SDS (1.6%) final concentration and incubated for 25 min at 65°C, 1400 rpm. The digested nuclei were transferred to a 50ml tube and 6.125ml of 1.15x ligation buffer was added. SDS inactivation was carried out by adding 375µl of Triton X-100 (1% final concentration) and samples were incubated for 1 hour at 37°C while shaking gently. Chromatin was then incubated with 100U T4 ligase, for 4 hours at 16°C followed by 30 min at RT. Proteinase K (300µg final) was added and DNA-protein complexes were uncross-linked (including undigested and digested controls) at 65°C O/N. DNA was recovered by adding RNase (300µg final) followed by phenol-chloroform extraction and ethanol precipitation. DNA concentration of the 3C template was determined and the 2nd digestion was carried out with NlaIII (Fermentas), 1unit/mg DNA and 17µl of 1x appropriate buffer), O/N. The enzyme was heat-inactivated and the DNA was recovered as before and ressuspended in 100µl milli-Q H_2O. Ligation was carried out in 14ml total volume of 1.15X with 200U of T4 ligase, for 4 hours at 16°C, plus 30 min at RT. DNA was recovered by phenol extraction and DNA purification. The primers used to verify digestion were: GAPDH-ApoIF 5'–GTATGACTGGGGGTGTTGGG-3'; GAPDH-ApoIR 5'–GGGGAGGCTCCTCCAGAATA-3'; ApoI-F 5'-AGGCATAAGGGGCTGACAAG-3'; SpeI-nested 5'–TAATCAGAGATTCAGATTTAAG-3'. The primers used for sequencing were BanII nested F3 5'– CACGTTTTTTGCCCTGAGCTTTG- 3' and BanII nested R3 5'-CTGGAGCCACCTTCAAGTTC-3'.

2.7 DNA sequencing

100ηg of purified DNA was mixed with 8ρmol of forward or reverse primer and 0.5µl of fluorescent dideoxy terminator nucleotides (BigDye, Applied Biosystems) in the appropriate reaction buffer.

The PCR program used for the sequencing reaction consisted of a denaturation step at 96ºC for 1 min followed by 25 cycles of 10 seconds at 96ºC, 5 seconds at 50ºC and 4 minutes at 60ºC. The PCR product was purified by using 2µl 3M Na-acetate, 2µl 125mM EDTA and 50µl absolute EtOH, spun at 3400 rpm for 30 minutes, washed with 70% cold EtOH at 3400 rpm for 15 minutes and dried at RT

for 5-10 minutes. DNA was ressuspended in 10µl Hi-Di Formamide (Applied Biosystems) and resolved by capillary electrophoresis on an Applied Biosystems 3730 DNA analyser (Applied Biosystems, USA).

Table 6. Primers used for sequencing

hGATA-1 F	5'-ATGGGTACCATGGAGTTCCCTGGCCTGGGG-3'
hGATA-1 R	5'-ATGGCGGCCGCTCATGAGCTGAGCGGAGCCAC-3'
pCDNA5ΔCMV UNIVF	5'-GCCGCAAAAAAGGGAATAAGGGC-3'
pCDNA5ΔCMV UNIVR	5'-CACCTACTCAGACAATGCGATGC-3'

2.8 Flow-activated cell sorting (FACS)

Peripheral blood mononuclear cells (PBMC) were pelleted at 1200 rpm, washed and ressuspended in 100µl RoboSep buffer (Stem Cell Technologies). The cells were incubated with the antibodies CD19-APC (BD Pharmingen), CD14-PE (eBioscience), CD3-efluor450 (eBioscience), and CD59-FITC (BD Pharmingen) or FLAER (Alexa 488 conjugated proaerolysin; Pinewood Scientific) for 30 minutes at 4ºC in the dark. After incubation cells were washed in PBS/2% FBS and sorted in a FACSAria II flow cytometer (BD Biosciences, USA). K652- and LBCL-GFP expressing cells were collected, washed in PBS/2% FBS and sorted. Dead cells were excluded by DAPI (Sigma) or 7-AAD (BD Pharmingen) staining. Analysis was performed using FlowJo software (Tree Star Inc., USA).

2.9 Generation of FRT stable cell lines

Flp-In-293 cells (Invitrogen) containing a single integrated FRT (Flp Recombination Target) site and stably expressing the lacZ-Zeocin fusion gene were co-transfected with pCDNA5-*PIGM*-GFP plasmid, in which the CMV promoter has been deleted, and the recombinase pOG44, as described in 2.16.
After transfection, cells were maintained under selection (100µg/ml hygromycin) for approximately 4 weeks. Single recombinant clones (see scheme below) were isolated by using cloning cylinders (Sigma), trypsinised and expanded in the appropriate medium. In order to confirm site-specific recombination, genomic

DNA was extracted using the GeneJET Genomic DNA Purification Kit (Fermentas) and PCR was carried out using the following 3 primers in the same reaction: F1 5'-CTATTCCAGAAGTAGTGAGG-3'; R1 5'-GTCGTCCATCACAGTTTGCC-3' and R2 5'-GATAGGTCACGTTGGTGTAG-3'. The PCR program consisted in an initial denaturation step of 5 minutes at 94ºC, followed by 35 cycles of 30 seconds at 94ºC, 45 seconds at 56ºC and 2 minutes at 72ºC and concluded with a final extension of 8 minutes at 72ºC. The PCR product was analysed on an agarose gel.

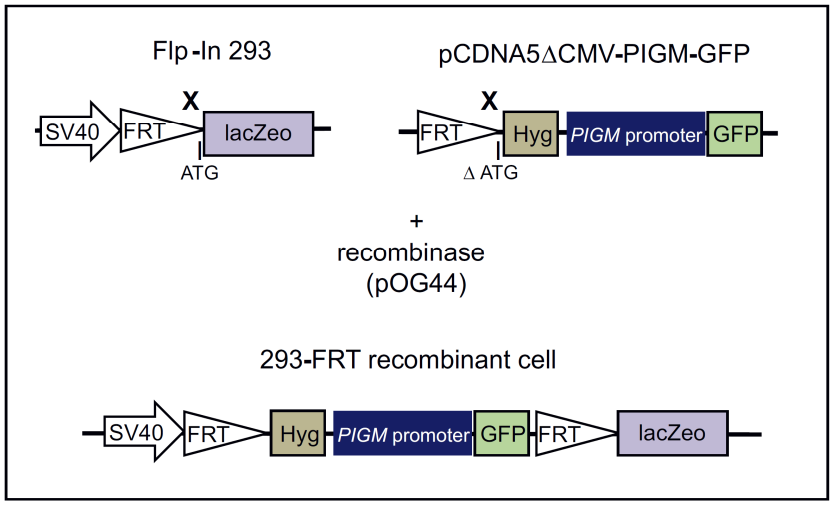


Fig. 10 – Schematic representation of the FRT system.

2.10 *In vitro* erythroid differentiation and cytospin staining

The two-phase erythroid culture system was adapted from Ohene-Abuakwa *et al*, 2005 (178) and Ronzoni, L. *et al*, 2008 (179).

PBMC ($2x10^6$ cells/ml) and $CD34^+$ cells ($1x10^6$ cells/ml) derived from cord blood were plated in erythroid medium, consisting of Stemspan (Stem Cell Technologies) supplemented with 100ηg/ml stem cell factor (SCF, Peprotech), 10ηg/ml interleukin-3 (IL-3) and 0.5 U/ml erythropoietin (EPO) during phase 1 (day0-7) of the cultures. After 7 days, $CD36^+$ cells were selected using anti-PE magnetic beads and the MiniMACS system (Miltenyi Biotec). CD36-PE cells

were re-plated in erythroid medium containing 4 U/ml EPO for 2-3 more days (Phase 2) and then sorted (for RNA) or magnetically separated (for ChIP).
Erythroid differentiation was monitored by flow cytometry analysis using the antibodies CD34-PerCP.Cy5.5 (Biolegend), CD36-APC (BD Pharmingen), CD71-PE (Biolegend) and GlyA-eFluor450 (eBioscience) and by May-Grünwald-Giemsa stained cytospin preparations (as described in 2.4).

2.11 Micrococcal nuclease (MNase) protection assay

$2x10^6$ cells were cross-linked with 1% paraformaldehyde in PBS, for 10 minutes at 37ºC. Cross-link was quenched by adding glycine to a final concentration of 125mM, and incubated on ice for 5 minutes. Cells were washed with ice-cold PBS and ressuspended in 1ml buffer A (25mM HEPES pH 7.4, 150mM NaCl, 0.1mM EDTA, 0.1% Triton-X 100, 10% glycerol, 1 mM DTT, 1x PI) followed by incubation for 15 minutes at 4ºC with rotation. The sample was split in two and centrifuged at 4ºC, 4500 rpm for 10 minutes to pellet nuclei. Nuclei were then ressuspended in 225µl of MNase buffer (25mM KCl, 12.5% glycerol, 50mM Tris pH 7.9, 10mM $CaCl_2$, 4mM $MgCl_2$, 1mM PMSF) and incubated with or without MNase (300units/ml) at the indicated concentration(s), at 37ºC for 15 minutes. Digestion was stopped by adding 25µl of MNase stop solution (10mM Tris pH 7.9, 200mM EDTA) and 250µl of digestion buffer (20mM Tris pH 7.9, 200mM NaCl, 50mM EDTA and 1% SDS). Cross-link was reverted at 65ºC overnight. The samples were treated with 6µg/ml RNase A (Fermentas) for 30 minutes at 37ºC followed by digestion with 50µg/ml Proteinase K (Fermentas), 30 minutes at 37ºC. DNA was recovered by phenol-chloroform extraction and ethanol precipitation. Micrococcal digestion was analysed in a 2% agarose gel and the optimal MNase concentration (147bp mononucleosome units) was used in subsequent experiments. Nucleosome occupancy was determined by RQ-PCR by calculating the ratio of digested to undigested chromatin for equivalent DNA amounts and using the delta Ct method for determination of the relative levels. The length of the *PIGM* promoter was covered by 100bp amplicons overlapping by 50bp. *GAPDH* promoter amplicon was used as an accessibility control.

Table 7. MNase protection assay primers

Designation	Forward primer	Reverse primer
i	5'-CTTCAACTCATCCTGTCCTG-3'	5'-GTAGGAGTGAAAGGAAGATAAG-3'
h	5'-GCGAGCTAGTGGAGTAAAGC-3'	5'-GGCTGCGCGCTACATCCTTC-3'
g	5'-TTTTTCACTCTCGGATCTCC-3'	5'-CCCCATTTCCAGTCCTCAG-3'
f	5'-CAGCTCGCTGCTCGTCGG-3'	5'-GGAGTGAGATTTCTTTGTAAGTG-3'
*e	5'-GGAAGAGGGCGGGGGGG-3'	5'-TCCGCTTGCCGGGAAATCc-3'
*d	5'-CACTCCATAAAATCCTCAAGCC-	5'-CCGCTTCTTCTTCCAGCGG-3'
c	5'-TCCCCTCCCACGGACCGG-3'	5'-AGCCGCCCGAGCCAAAAAC-3'
*b	5'-CGAGACCGTCCATCCAGAGG-3'	5'-CATGAAGCCCCGCCCCCG-3'
*a	5'-GACCGCGCGGGGCTGGAG-3'	5'-CTCCCACCCGCCAGGCTG-3'
1	5'-CGGATGTGATAGTCTGCAGTC-3'	5'-CCTTCGTACTTCTAACCTTCCC-3'
2	5'-TGCGGCGGCCACCTGCTG-3'	5'-GCCCATGATCTGACCGTGCG-3'
GAPDH	5'-TACTAGCGGTTTTACGGGCG-3'	5'-TCGAACAGGAGGAGCAGAGAGCGA -3'

Tm=60ºC *Tm=63ºC

2.12 Plasmids and cloning

A PCR-based technique was previously used to clone a 2Kb fragment (-2000/+1) of the human *PIGM* promoter from either WT or MUT gDNA of lymphoplastoid cell lines into the pGL3-Basic vector (Invitrogen) using the KpnI and NcoI restriction sites. A pair of truncated constructs (-910WT and -910MUT) of the 5'end was then generated by PCR amplification and monitored on agarose gel.
The 910bp fragment (-910/+1) was digested with KpnI and NcoI (Fermentas) and sub-cloned into the pCDNA5/FRT vector that contains a GFP reporter gene that is expressed under the activity of the promoter in study. The vector was linearized with KpnI and dephosphorylated with SAP (Fermentas). Both vector and insert were incubated with Klenow (Fermentas) and 1µl 10mM dNTPs for 15 minutes at RT and then ligated.

hGATA-1 cDNA (1241bp fragment) previously cloned in the pREP4 vector was digested using KpnI and NotI restriction enzymes (Fermentas) and the MigR1-ratCD2 vector (kindly donated by G.Bohn) was linearised with EcoRI. Both vector and insert were incubated with Klenow and 1µl 10mM dNTPs for 15 minutes at RT and the ligated.
Ligation reactions containing 1:6 (fragment:vector) molar ratio were performed by using T4 DNA ligase (Fermentas) in a final volume of 10µl, O/N at 4ºC .

Sp1-DNA binding domain (DBD) was amplified by PCR from pEBGN-Sp1 (kindly donated by Dr.Gerald Thiel) using BglII and EcoRI restriction sites. The fragment was sub-cloned into the MigR1-GFP plasmid and ligation was carried out as described above.

5µl of each ligation were used to transform chemically competent cells (as described in 2.3). All plasmids were verified by restriction enzyme digestion and monitored in an agarose gel. Plasmids were also verified by direct sequencing (see Tables 6 and 8).

Scramble shRNA, GATA-1 shRNA and Sp1 shRNA oligonucleotides were designed in the Mission shRNA web tool (http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/shrna.html) and purchased from Sigma. Oligonucleotides were phosphorylated with PNK (Fermentas) for 1 hour at 37ºC followed by annealing for 5 minutes at 95ºC and slowly cooled to room temperature (RT). Annealed oligonucleotides were then ligated into the lentiviral plasmid pLKO.1-GFP previously digested with AgeI and EcoRI restriction enzymes (Fermentas). Ligation reaction was carried out by T4 DNA ligase, O/N at 4ºC and 5µl of the ligation reaction were used to transform chemically competent cells (as described in 2.3). Plasmids were verified by restriction enzyme digestion using EcoRI and NcoI (see Appendix A, Fig. A6).

Table 8. Primers used for cloning

-910 KpnI F	5'-CTAATGGTACCAAATACAGAGATACTGAGGCAT-3'
+1 NcoI R	5'-TAATCCATGGATCTGACCGTGCGACAGCTGC-3'
DNSp1BglII F	5'-ATAATAGATCTATGCCTCCTAAAAAGAAGCGCAAGGTAG-3'
DNSp1EcoRI R	5'-ATAATGAATTCATGCCTCCTAAAAAGAAGCGCAAGGTAG-3'
shRNA GATA1 F1	5'-CCGGCGGCCCAAGAAGCGCCTGATTCTCGAGAATCAGGCGCTTCTTGGGCCGTTTTT-3'
shRNA GATA1 R1	5'-AATTCAAAAACGGCCCAAGAAGCGCCTGATTCTCGAGAATCAGGCGCTTCTTGGGCCG-3'
shRNA Sp1 F1	5'-CCGGGCAGCAACTTGCAGCAGAATTCTCGAGAATTCTGCTGCAAGTTGCTGCTTTTTG-3'
shRNA Sp1 R1	5'-AATTCAAAAAGCAGCAACTTGCAGCAGAATTCTCGAGAATTCTGCTGCAAGTTGCTGC-3'
shRNA Sp1 F2	5'-CCGGCCCAAGTTTATTTCTCTCTTACTCGAGTAAGAGAGAGAAATAAACTTGGGTTTTT-3'
shRNA Sp1 R2	5'-AATTCAAAAACCCAAGTTTATTTCTCTCTTACTCGAGTAAGAGAGAAATAAACTTGG-3'

2.13 Rapid amplification of cDNA ends (RACE)

The 5'-RACE analysis was carried out using the ExactSTART Eukaryotic mRNA 5'- & 3'-RACE kit (Epicentre). 2ug of total RNA was treated with Apex Heat-Labile alkaline phosphatase to remove the 5'-phosphates of truncated mRNA, at 37ºC for 15 minutes. After extraction by phenol-chloroform and ethanol purification, RNA was treated with tobacco acid pyrophosphatase (TAP) to remove the 5'-cap structure from intact, full-length mRNA, at 37ºC for 30 minutes. 5'-RACE acceptor oligo was ligated to the 5'end of the mRNA by adding 1µl TAP stop buffer, 1µl 2mM ATP solution and 1µl T4 RNA Ligase, at 37ºC for 30 minutes. 3'-polyadenylated RNAs were then reverse transcribed into first-strand cDNA by adding 1µl cDNA synthesis primer (oligo d(T) sequence with a PCR priming site sequence at its 5' end), 2µl dNTP, 2µl MMLV RT Buffer and 1µl MMLV reverse transcriptase. The reaction was incubated at 37ºC for 1 hour followed by 10 minutes at 85 ºC to inactivate the enzyme. Second-strand cDNA was synthesised by adding 5µl PCR primer 1 (which anneals to the PCR priming site at the 5'-RACE acceptor oligo), 5µl oligo d(T) (instead of PCR primer2) and 2.5U of DNA polymerase enzyme with the appropriate buffer following the conditions: 95ºC for 30 seconds (initial denaturation); 95ºC for 20 seconds, 60ºC for 20 seconds and 72ºC for 3 minutes, repeated for 21 cycles. The PCR reactions were carried out by using PCR Primer 1 and β-actin reverse primer (5'-AGGTGTGGTGCCAGATTTTC-3') or PIGM reverse primer (5'-CTTCTAACCTTCCCTTCGGTTCTTT–3').

Both PCR products were analysed on a 2% agarose gel. The PIGM PCR product was then cloned into the pJET- vector according to the manufacturer's instructions. Following bacterial transformation, the plasmid DNA was sequenced.

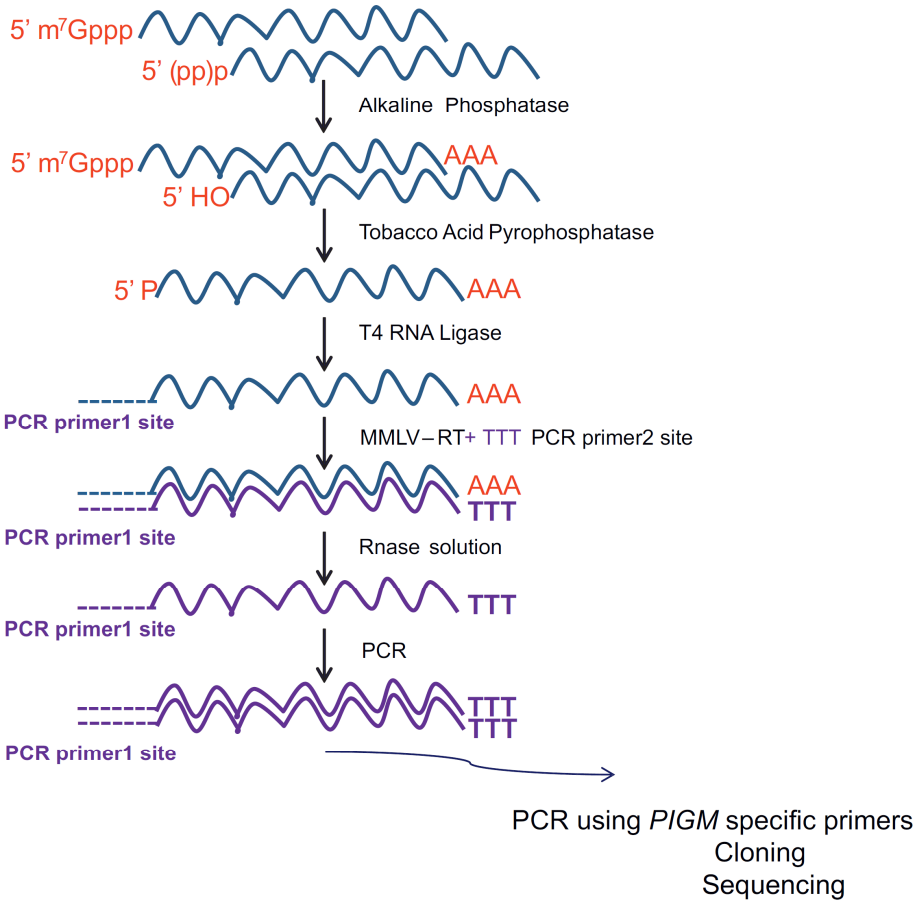


Fig. 11 – Schematic representation of RACE.

2.14 Reporter assays

Adherent cells were seeded in a 24 well plate at a concentration of $1.5x10^5$-$2x10^5$cells/well and transfected with 450ηg of plasmid DNA expressing the luciferase reporter gene and 50ηg of pRL (renilla) used for normalisation (see 2.16).

Suspension cells were transfected with 1.8µg luciferase reporter gene and 0.2 µg of renilla by electroporation.

48 hours after transfection cells were washed with PBS and lysed in 100ul of 1X Passive Lysis Buffer (PLB; Promega) under agitation at 300rpm for 15 minutes at RT. Cell lysate was collected and spun at 14.000rpm for 5 minutes at 4ºC, to

remove cell debris. 20µl of the cell lysate were used to measure luciferase activity by using the Dual-luciferase reporter assay system (Promega), on a Fluoroskan Ascent Luminometer (Thermo Scientific), for 100ms. Measurements were performed in triplicate and luciferase reporter activity was normalised with respect to renilla.

2.15 RNA extraction, cDNA synthesis and quantitative reverse transcriptase-PCR (qRT-PCR)

Total RNA from cell lines was extracted using the GeneJet RNA purification kit (Fermentas) and total RNA from primary cells in small number was extracted using the NucleoSpin RNA XS kit (Macherey-Nagel). When appropriated, gDNA was digested by DNase treatment (1U/ug RNA; Fermentas). RNA was reverse-transcribed into cDNA in a thermocycler using the RevertAid First strand cDNA synthesis kit (Fermentas). Briefly, 100 to 500ng of RNA was incubated with 1µl oligo d(T) primers at 65 ºC for 5 minutes and after cooled down a reaction mix containing 2µl 10mM dNTPS, 200 units Revert Aid, and 20 units RNAse inhibitor (Ribolock) was added and samples were left at 42ºC for 60 min. cDNA was diluted according to the initial amount of RNA and gene expression was measured with specific primers and Maxima POWER SYBR Green master mix (Fermentas) in a AB 7500 Real-time PCR system. The PCR program is described above (see 2.5). All primers were first tested for linearity by using serial dilutions of cDNA. Single PCR product was verified by melting curve dissociation and expression values were obtained by using the Δ Ct method ($2^{-(\text{Ct gene-Ct reference})}$).

Table 9. qRT-PCR primers

Designation	Forward primer	Reverse primer
RT-GATA-1	5'- ACACCAGGTGAACCGGCCAC-3'	5'- CCTTCGGCTGCTCCTGTGCC-3'
RT-γ EX2	5'- TGGCAAGAAGGTGCTGACTTC-3'	5'-GCAAAGGTGCCCTTGAGATC-3'
RT-γ EX3	5'-AATCCATTTCGGCAAAGAATTC-3'	5'-CCACTCCAGTCACCATCTTCTG-3'
RT-PIGM	5'- CAGAGTCGCCCTGGTTTTCT-3'	5'- TTCCAAAGAGCTCGCTGAGG-3'
RT-PIGX	5'–CTTCCCATAACCCTCCACCT-3'	5'-GGAGCTCGTACAAACAAGCC-3'
RT-Sp1	5'-CTGGTCCGCCCTCTGACCAAGA-3'	5'-TCGAGCCTGTGAAAAGGCACCA-3'
GFP	5'-GTTCATCTGCACCACCGGC-3'	5'-CGGGTCTTGTAGTTGCCGTC-3'
GAPDH	5'-GAAATCCCATCACCATCTTCCAGG-3'	5'-GAGCCCCAGCCTTCTCCATG-3'
DHFR	5'-GAGCTCGAGCCCAAGGGATA-3'	5'-ATGAGCTCCTTGTGGAGGTTC-3'
α-globin	5'-GAGGCCCTGGAGAGGATGTTCC-3'	5'-ACAGCGCGTTGGGCATGTCGTC-3'

2.16 Transfections

In order to obtain 293 *PIGM* promoter-GFP cells, Flp-In-293 cells were seeded in 10cm^2 plates (2x10^{6}cells/plate). On the day after, cells were co-transfected with 1μg of pCDNA5-ΔCMV-FRT containing the *PIGM* promoter and 9μg of the Flp recombinase expression plasmid, pOG44 (Invitrogen) using the $CaCl_2$ method. Briefly, the plasmid solution was made up to a final volume of 450μl with low TE buffer and 50μl of 2.5M $CaCl_2$ was added to the plasmid mixture. Following 5 minutes incubation, 500μl of 2x HBS was added drop wise to the 500μl DNA-TE-$CaCl_2$ complex while vortexing at full speed. Immediately after vortexing, the precipitate was added drop wise to the 293T cells. After 14-16 hours, cell media was replaced.

HeLa, 293 cells and 293 *PIGM* promoter-GFP clones were seeded in 6-well plates at a concentration of 1.5x10^5 cells/well and 2.5x10^5 cells/well, respectively, and transfected with plasmid DNA (typically 2μg) by lipofection (Lipofectamine 2000, Invitrogen). 4 hours after transfection, cell media was replaced.

K562 cells were transfected by electroporation in 0.4cm cuvettes, using a Gene Pulser II Electroporation System (BioRad ,USA).

2.17 Viral production and transduction

Lentiviruses were generated by co-transfection of $3x10^6$ 293T cells with plasmids encoding gag-pol (5µg pMDLg/pRRE), Rev (2.5µg of pRSV-REV), envelope plasmid (3µg VSV-G) and shRNA-pLKO.1 GFP (12µg). For retroviral production, cells were co-transfected with gag-pol (12µg pMDLg/pRRE), envelope plasmid (3µg VSV-G) and MigR1GFP or hGATA-1 (10µg). The plasmid solution was made up to a final volume of 450µl with low TE buffer. Finally, 50µl of 2.5M $CaCl_2$ was added to the plasmid mixture. Following 5 minutes incubation of the plasmid mixture, 500µl of 2x HBS was added dropwise to the 500µl DNA-TE-$CaCl_2$ complex while vortexing at full speed. Immediately after vortexing, the precipitate was added dropwise to the 293T cells. After 14-16 hours, cell media was replaced and cell supernatants were collected 24 and 48 hours after changing the media and spun for 2 hours at 23000 rpm in a ultracentrifuge. Pellet was ressuspended in culture medium and stored at –80°C. Before freezing, viral supernatant was taken to infect 293T cells in order to determine viral titre (see Appendix A, Fig. A7).

K562 were transduced using a spin-infection method and polybrene (8ug/ml final concentration). The cells were spun down for 1 hour at 2000 rpm, 32ºC and then kept at 37ºC overnight. Medium was changed 12 hours thereafter.

2.18 Western-blot

Cells were lysed with RIPA lysis buffer (buffer composition) and protein concentration was determined by employing the method of Bradford, following the kit manufacturer's instructions (BioRad). BSA standards (BioRad) were used to generate a standard curve. Typically, 25µg of whole cell protein extracts were size-fractionated by electrophoresis on 12% SDS-polyacrilamide gel and transferred onto PVDF membranes (BioRad) using the Trans-Blot SD Semi-Dry Transfer system (BioRad) with constant voltage of 12V for 45 min. Membranes were blocked for 1h30 at RT with PBS/0.1% Tween20 containing 5% non-fat dry milk and then washed three times with PBS/0.1% Tween20.

Specific antibodies for GATA1 (ab11852, abcam), GFP (2555, Cell Signaling), and β-actin (sc-1616, Santa Cruz) were diluted in the appropriate solution and

used to detect the protein of interest, O/N at 4ºC. Next, the membranes were washed three times with PBS/0.1% Tween20 before incubation for 2 hours at RT with the HRP-conjugated secondary antibody (Santa Cruz). ECL (enhanced chemiluminescence) solution (Amersham) was used for signal detection. ImageJ software was used to determine % of protein expression.

2.19 Statistical analysis

Results are expressed as mean ± standard error of the mean (SEM) unless otherwise stated. Data was analysed using the unpaired two-tailed student's t test or one-way ANOVA with Bonferroni correction for multi-comparison test (GraphPad Prism5, CA, USA). P values < 0.05 were considered statistically significant.

Chapter 3 – Results

3.1 Elucidating the role of *PIGM* transcription in *PIGM*-associated IGD

3.1.1 Introduction

In *PIGM*-associated IGD, the expression of GPI at the surface of haematopoietic cells varies in a cell-type specific manner. The differential expression of GPI is observed not only between different haematopoietic lineages but also between cells belonging to the same lineage. Regarding the lymphoid lineage, as I will show later, the majority of B lymphocytes are mostly GPI negative (GPI-) with a small proportion of cells being GPI positive (GPI+), whereas the T lymphocytes are mostly GPI+. In striking contrast with PNH, the erythrocytes are mostly GPI+ (Fig. 12), hence the absence of significant haemolysis in patients (87).

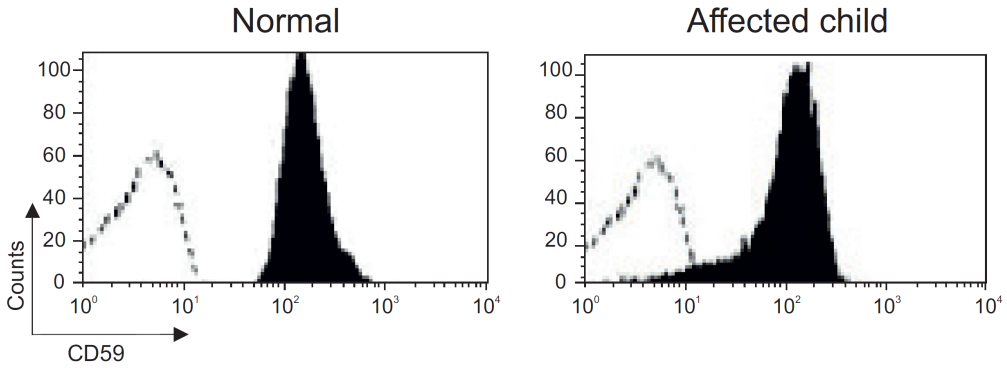


Fig. 12 – Flow cytometric profile of the red blood cells (RBC) from a normal donor and affected child (-270C>G homozygous) stained with the GPI-linked antigen CD59.

Although the differential expression of GPI has been linked to the -270C>G mutation in the regulatory region of *PIGM* (87), **a direct correlation between**

***PIGM* mRNA levels and expression of GPI in patient cells has not been shown.** Additionally, **the contribution of PIGX, the component of the GPI MT-I that stabilises PIGM, has not been investigated.**

It is known that gene expression is influenced by nucleosome positioning at functional regulatory regions such as promoters. The underlying DNA sequence and binding of generic and tissue-specific TFs constitute some of the elements that determine the occupancy of the nucleosomes, which therefore affect chromatin accessibility and transcription output.
Another important aspect in gene regulation is the presence of a single or multiple TSSs, with multiple TSSs often associated with highly GC-rich promoters (21) like that of *PIGM*.

Characterising the architectural organization of the *PIGM* promoter in terms of nucleosome occupancy and determining the position of the TSS(s) in cells from different haematopoietic lineages might be relevant for better understanding of the regulatory mechanism(s) driving *PIGM* transcription.

The aims of this chapter are to investigate **a)** *PIGM* and *PIGX* mRNA expression in normal haematopoietic cells **b)** the correlation between *PIGM* mRNA levels and GPI in cells from patients with *PIGM*-associated IGD, **c)** the impact of the -270C>G mutation on the promoter activity of *PIGM* in erythroid cells, **d)** the promoter accessibility in terms of nucleosome occupancy and TSS(s) positioning.

3.1.2 Experimental design

Gene expression was evaluated in cell lines and primary haematopoietic cells by qRT-PCR. For this purpose, primary haematopoietic cells from peripheral blood samples were obtained from normal donors and patients with IGD and isolated by FACS sorting. After isolation, cells were assayed for mRNA expression or maintained under appropriate culture conditions for induction of erythroid differentiation.
PIGM-promoter activity was studied by employing luciferase reporter assays.

The nucleosome occupancy was determined by micrococcal nuclease protection assay and the positioning of the TSS was resolved by 5'-RACE.

3.1.3 Results

The -270C>G mutation directly abrogates *PIGM* transcription in a cell-type specific manner

In order to gain insight on the degree of variability of *PIGM* expression, *PIGM* mRNA levels were first evaluated in haematopoietic cells derived from peripheral blood of normal donors.

B lymphocytes, T lymphocytes and monocytes were obtained by FACS sorting after staining of peripheral blood mononuclear cells (PBMC) with the appropriate lineage-specific antibodies (CD19, CD3 and CD14, respectively) (Fig. 13A) while neutrophils were isolated after lysis of the red blood cells pellet derived from Ficoll centrifugation. The typical morphology of neutrophils was confirmed by MGG staining (Fig. 13C).

A unimodal pattern of GPI expression was seen in B lymphocytes, T lymphocytes and monocytes obtained from normal donors, as assessed by FLAER staining (Fig. 13B). FLAER is a fluorescein-labeled modified proaerolysin that binds to GPI (180).

While GPI expression in mature red blood cells is nearly normal, whether erythoid precursors in IGD also express high levels of GPI was not investigated. To address this, I generated erythroid precursors using a method suitable for *in vitro* generation of nucleated erythroid precursor cells from PBMC (see Chapter 2, 2.10). PBMC constitute one of the sources of $CD34^+$ progenitor cells that in the appropriate culture conditions have the capacity to commit into erythroid progenitor cells. Given that *in vitro* erythroid differentiation slightly varies from sample to sample, I describe below a representative example of a normal culture. Erythroid differentiation was monitored by FACS analysis of the cell surface markers CD36, CD71 and GlyA (see Fig. 8) at days 0, 7 and 10 of the liquid culture (Fig. 14). At day 0, 18.9% of cells were $CD36^+/CD71^-/GlyA^-$ and after 7

days in the appropriate erythroid culture conditions, $CD36^+$ cells expanded to 25.9%. At this stage, almost all $CD36^+$ cells expressed the non-erythroid specific transferrin receptor, CD71, with 70.3% also expressing GlyA, an erythroid-specific marker of late precursor and mature cells. By re-plating the $CD36^+$ cells for an additional period of three more days in Phase II medium, all cells became $CD71^+$ and gained GlyA expression (89.5%). The stage of maturation of the erythroid precursor cells corresponded to the morphological modifications as assessed by microscopic examination of MGG stained cytospins (Fig. 14). Accordingly, at day 7, erythroid precursor cells presented a large nucleus, characteristic of basophilic erythroblasts, and later in differentiation the nucleus was reduced in size and chromatin became more condensed. These cells are called polychromatic-erythroblasts.

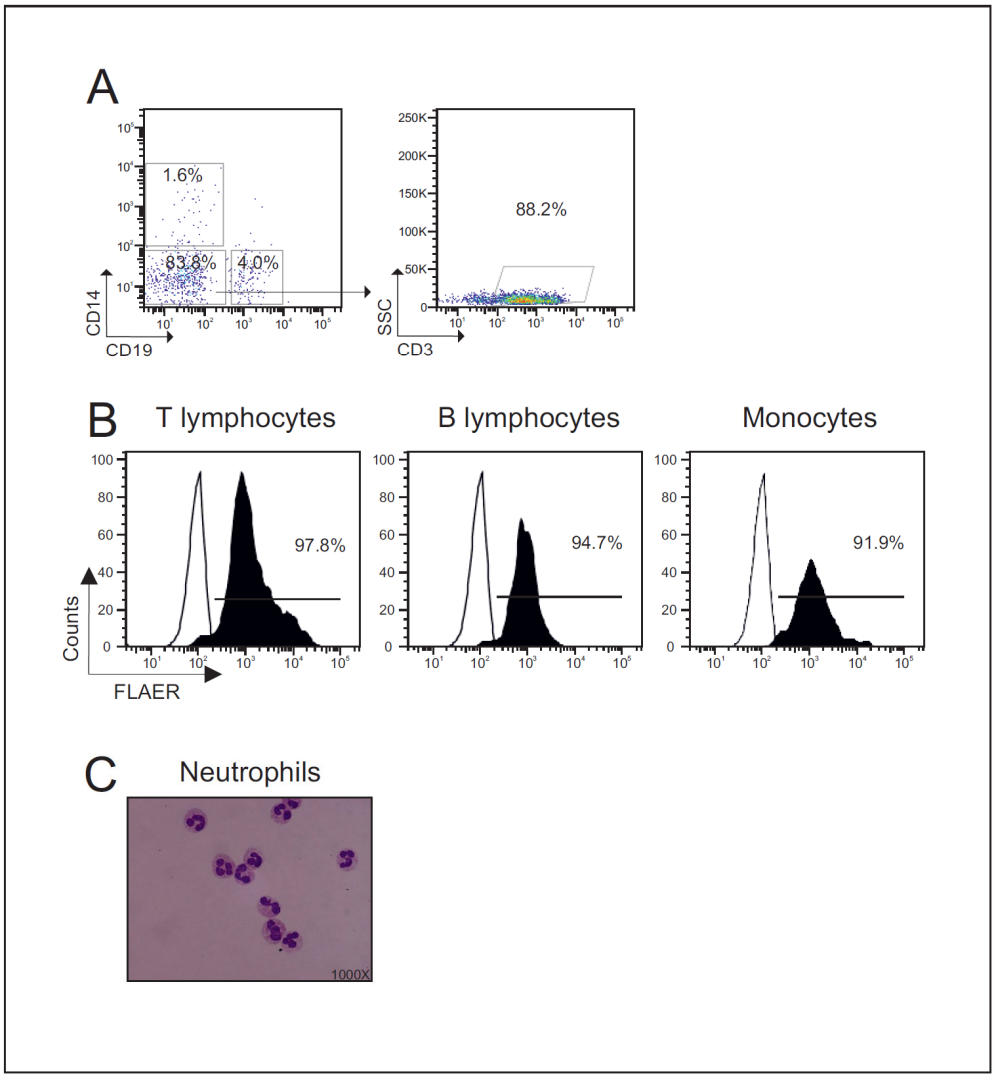


Fig. 13 – Isolation of the haematopoietic cells derived from peripheral blood of a normal donor and assessment of GPI expression. Peripheral blood mononuclear cells were identified according to physical properties SSC (side scatter) vs. FSC (forward scatter) and doublets were excluded by double gating on FSC-W (width) and FSC-A (area). **A** – Cells were stained with CD19, CD3 and CD14 antibodies. T lymphocytes ($CD3^+$) were gated on the double negative $CD19^-CD14^-$ fraction; B lymphocytes corresponded to $CD19^+$ cells and monocytes corresponded to $CD14^+$ cells. **B** - Flow cytometric profile of T lymphocytes, B lymphocytes and monocytes stained with FLAER. The unimodal pattern of expression was observed in all cell types, with the majority being GPI^+ as compared to unstained control. Numbers above the gates indicate the frequency of positively stained cells. **C -** Neutrophils were obtained by RBC lysis and morphology, characterised by a multi-lobed nucleus, was assessed by May-Grunwald-Giemsa (MGG) staining.

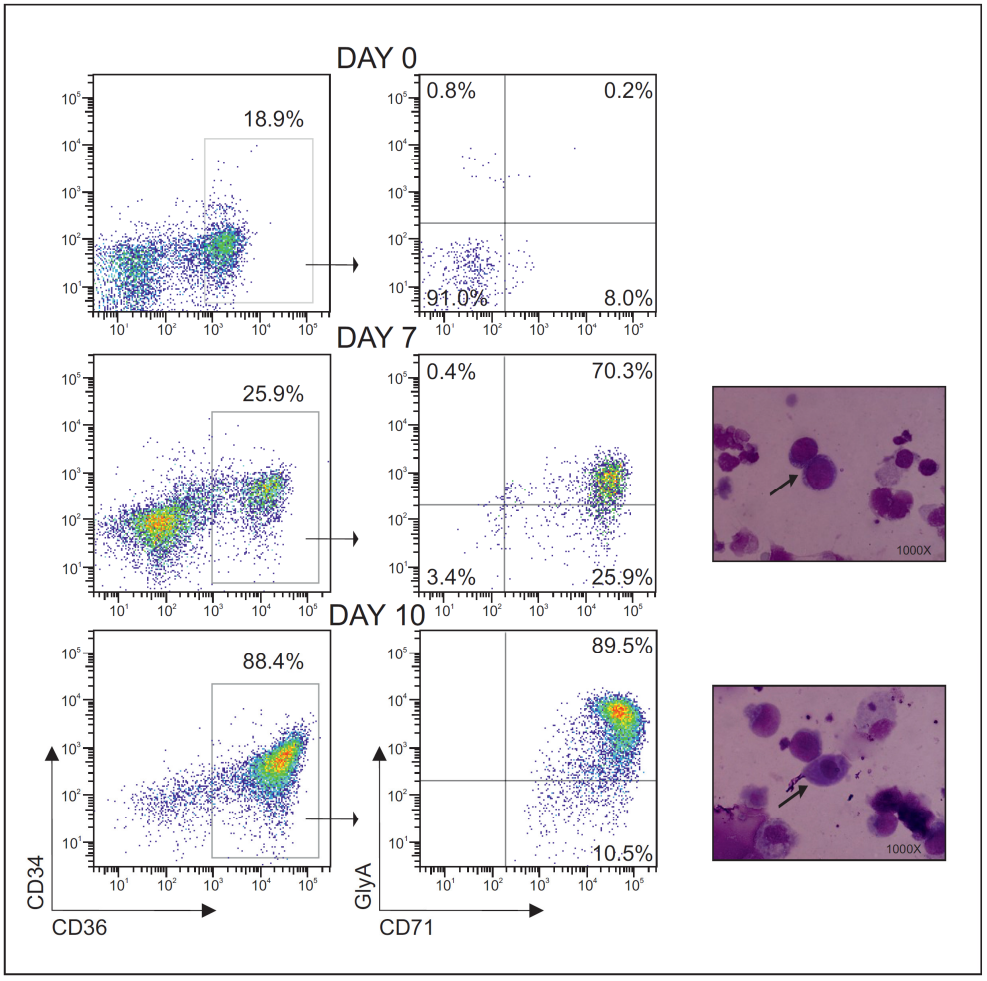


Fig. 14 - *In vitro* erythroid differentiation from a peripheral blood sample. Peripheral blood mononuclear cells were identified according to physical properties SSC (side scatter) vs. FSC (forward scatter). Doublets were excluded by double gating on FSC-W (width) and FSC-A (area) and dead cells were excluded by DAPI staining. Erythroid differentiation was monitored by flow cytometry analysis using CD34, CD36, CD71 and GlyA antibodies. CD71 and GlyA expression were assessed on $CD36^{+}$ cells. Erythroid precursor cells were isolated by magnetic separation after staining with CD36-PE antibody and incubation with anti-PE beads. Morphologic characteristics of selected cells were assessed by MGG staining. Cytospins are shown to illustrate the morphologic changes as erythroid cells mature from basophilic erythroblasts with a large nucleus (Day 7) through polychromatic-erythroblasts with a nucleus reduced in size accompanied by condensation of chromatin (Day10).

After isolation of the various haematopoietic cell types from samples collected from different donors, *PIGM* mRNA levels were evaluated. *PIGM* expression was variable amongst different cell types. *PIGM* mRNA levels were highest in

neutrophils followed by T lymphocytes and B lymphocytes with monocytes expressing slightly less *PIGM* mRNA. In purified $CD36^+$ erythroid precursor cells, *PIGM* levels decreased as cells underwent differentiation. At day 7, mRNA levels were comparable with those in the monocytes and later in differentiation, at day 10, decayed approximately by 2.5-fold (Fig. 15).

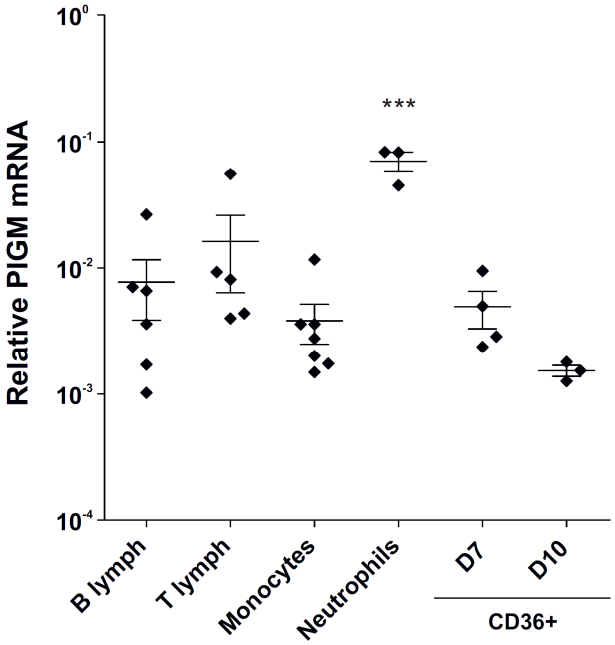


Fig. 15 - ***PIGM*** **mRNA expression in haematopoietic primary cells derived from normal donors.** B lymphocytes (B lymph), T lymphocytes (T lymph) and monocytes were isolated after staining with CD19, CD3 and CD14, respectively (see Fig. 13A). Neutrophils were isolated by lysis of the red cells (see Fig. 13C) and $CD36^+$ erythroid precursor cells (see Fig. 14) were isolated by magnetic separation. B lymphocytes: n=7, T lymphocytes: n=5, Monocytes: n=7, Neutrophils: n=3, Ery Day 7 n=4, Ery Day 10 n=3. Expression values were normalised to *GAPDH* and results are presented as mean ± SEM, *** $p<0.001$.

The reduction in the mRNA levels of *PIGM* seen during erythropoiesis was in agreement with RNA-seq data obtained from mouse fetal liver erythroid precursor cells isolated at different stages of differentiation (154). As shown in Fig. 16, the number of reads mapping to *PIGM* and to the erythroid-specific gene *GATA-1*, were reduced in later stages of fetal erythropoiesis (R4 and R5).

Whether changes at the mRNA expression levels mirrored changes at the protein level in these cells could not be assessed due to the lack of good quality anti-PIGM antibody.

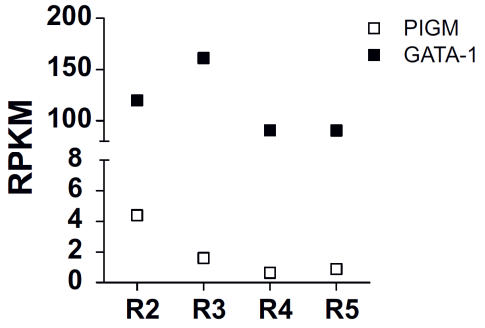


Fig. 16 – RNA-seq analysis in mouse fetal liver erythroid precursor cells. Fetal liver cells were double-labeled for erythroid-specific TER119 (GlyA equivalent in humans) and non erythroid-specific transferrin receptor (CD71) and then sorted by flow-cytometry. R2 – proerythroblasts and early basophilic erythroblasts, R3 – early and late basophilic erythroblasts, R4 – polychromatophilic and orthochromatophilic erythroblasts and R5 – late orthochromatophilic erythroblasts and reticulocytes. Gene expression values were calculated as Reads Per Kilobase of Exon Model Per Million Mapped Reads (RPKM); RPKM values from R3, R4 and R5 were normalised with respect to R2 and expressed as a fold change relative to R2 (154). Data was retrieved from the Gene Expression Omnibus database under the accession number GSE27893.

Having seen that mRNA levels of the housekeeping gene *PIGM* are variable in haematopoietic cells of normal donors, the next step was to establish in haematopoietic cells from IGD patients the relationship between *PIGM* mRNA levels and cell surface GPI expression.

For this purpose, PBMC of an affected child, collected before NaBu treatment and cryopreserved for several years, were used. Accordingly, cells from 2C (homozygous MUT from Family 2, see Fig. 9) were stained with antibodies against CD19, CD3, CD14 and FLAER and then the haematopoietic cell populations were examined according to expression of GPI into two fractions: GPI positive (GPI+) and GPI negative (GPI-). For the purpose of direct comparison, cryopreserved PBMC from the father of the affected child, 2E (heterozygous for the -270 substitution), were also analysed. As also described in (87), I found that haematopoietic cells from heterozygous individuals presented

normal levels of cell surface GPI. In cells from the affected child, the B lymphocytes showed a bimodal distribution in which the majority of the cells (70.9%) were GPI negative while a small proportion (29.1%) was GPI positive. Normal GPI expression was seen in 89% of the T lymphocytes with the remainder (11%) being GPI negative. GPI expression on monocytes in both homozygous and heterozygous individuals was unimodal and normal, with all cells staining with FLAER (Fig. 17).

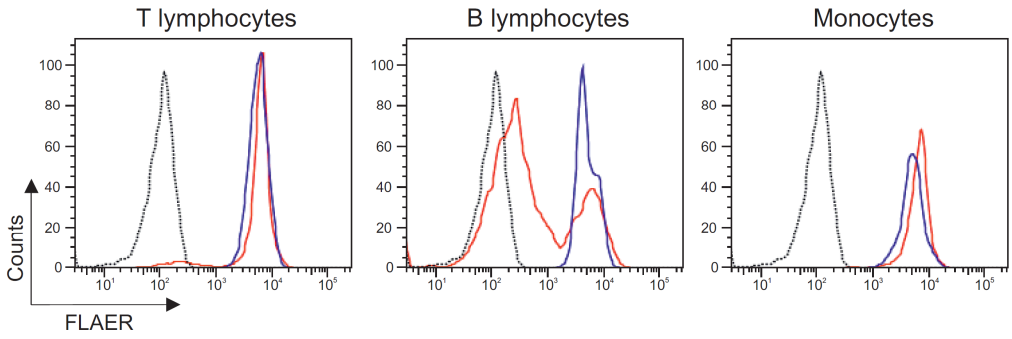


Fig. 17 – GPI expression profile of haematopoietic cells isolated from individuals of Family 2. Histograms show GPI expression as assessed by FLAER. The gating strategy used to identify $CD3^+$ T lymphocytes (left), $CD19^+$ B lymphocytes (middle) and $CD14^+$ monocytes (right) was described above (Fig. 13A). Dotted line – unstained; blue solid line – heterozygous individual 2E; red solid line – homozygous MUT individual 2C.

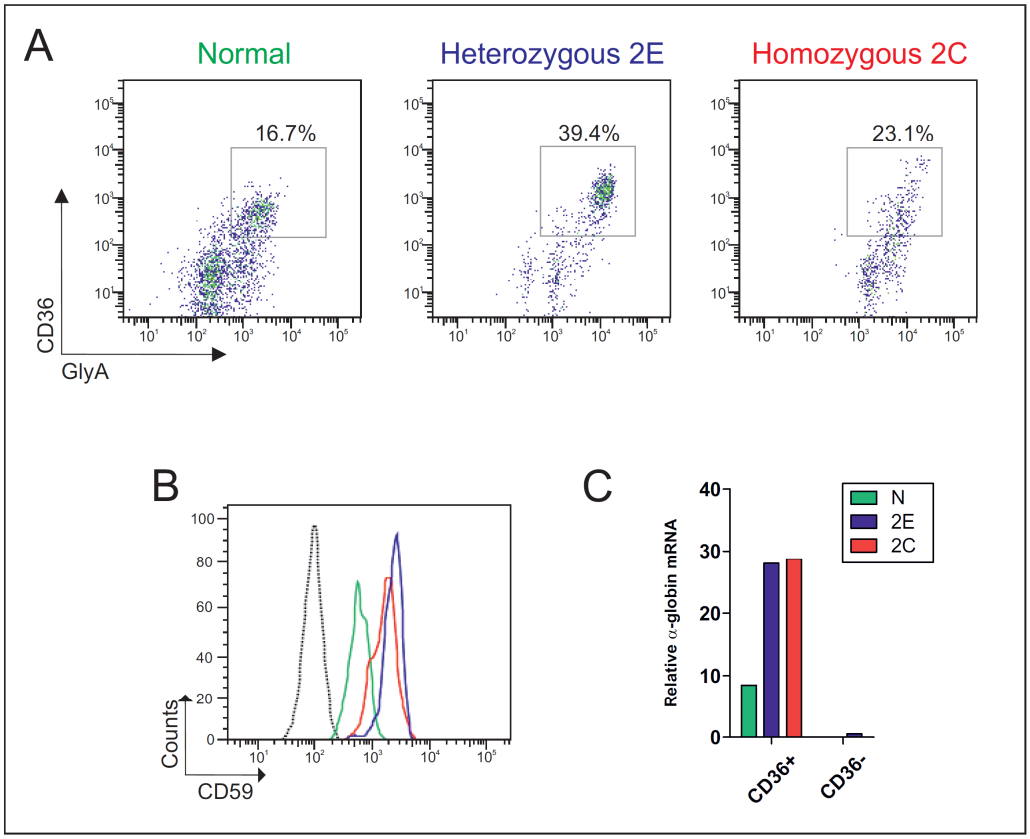


Fig. 18 – *In vitro* erythroid differentiation from PBMC of individuals from Family 2 and a normal control, at day 7. A - Erythroid differentiation was monitored by flow cytometry analysis using CD34 (not shown), CD36 and GlyA antibodies. Left – normal donor; middle – heterozygous individual, 2E; right – homozygous MUT individual, 2C. **B –** Flow cytometric profile of the GPI-linked antigen CD59. Dotted line – unstained; green solid line – normal donor; blue solid line – 2E; red solid line – 2C. Double positive $CD36^{+}/GlyA^{+}$ erythroid precursor cells were sorted by flow cytometry. **C -** $CD36^{+}$ erythroid precursor cells and $CD36^{-}$ control cells were tested for α-globin expression by qRT-PCR. Expression values were normalised to *GAPDH*. n=1

Erythroid precursor cells from individuals 2E and 2C were also generated *in vitro* from cryopreserved PBMC, following the approach described above. Briefly, erythroid differentiation was monitored by analysis of CD34, CD36 and glycophorin A (GlyA) expression. The last constitutes a specific marker of late, mature erythroid cells. To further characterise these cells, expression of the erythroid-specific α-globin gene was also assayed by qRT-PCR in WT,

heterozygous (2E) and homozygous MUT (2C) $CD36^+$ cells at day 7 of differentiation (Fig. 18).

In 2C and 2E, as assessed by GlyA expression, a higher frequency of $CD36^+$ cells were more mature than the $CD36^+$ cells derived from a normal donor (Fig. 18A). This was consistent with the 3-fold higher expression levels of α-globin in total $CD36^+$ cells from 2E and $CD36^+$ 2C than in normal donor $CD36^+$ cells. As expected, α-globin was not expressed in $CD36^-$ cells (Fig. 18C).

In order to evaluate whether the -270C>G erythroid precursor cells (2C) expressed normal levels of GPI, staining with the GPI-linked antigen CD59 was carried out. Contrasting with the bi- or tri-modal pattern seen in PNH erythroid cells (84), the result showed a unimodal pattern of CD59 expression in $CD36^+$GlyA+ cells from 2E and 2C, interestingly higher than in $CD36^+$GlyA+ normal cells. Staining with FLAER was not tested since it cannot bind to erythrocytes or platelets.

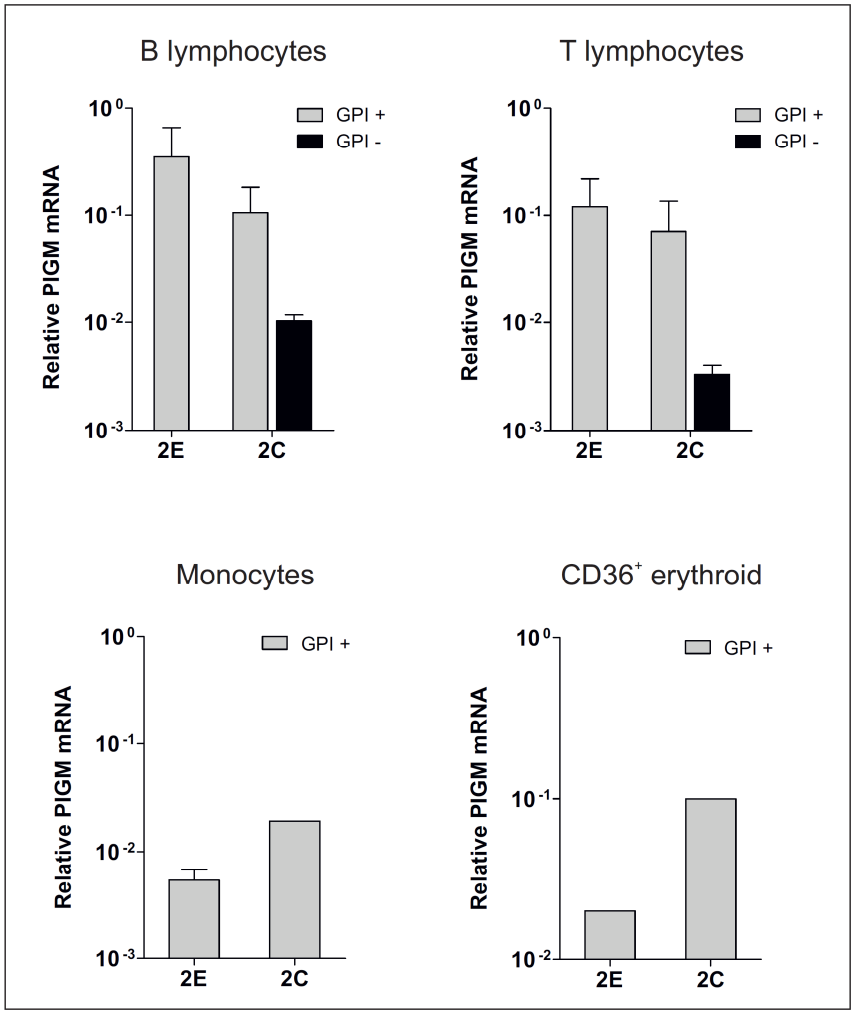


Fig. 19 - *PIGM* mRNA expression in haematopoietic cells isolated from PBMC of individuals from Family 2. The gating strategy used to identify $CD19^+$ B lymphocytes, $CD3^+$ T lymphocytes and $CD14^+$ monocytes was described above (see Fig. 13A). $CD36^+$ erythroid precursor cells were generated *in vitro* from PBMC and collected at day 7 (see Fig. 14). Haematopoietic cells were sorted according to expression of GPI into two fractions: GPI negative (GPI-) and GPI positive (GPI+). 2E – heterozygous individual; 2C – homozygous MUT individual. When samples were available, *PIGM* mRNA expression was assessed in two biological replicates. Expression values were normalised with respect to *GAPDH*.

After isolation of the cells belonging to the lymphoid, myeloid and erythroid haematopoietic lineages, *PIGM* mRNA expression was measured (Fig. 19). Regarding the B lymphocytes of the affected child (2C), *PIGM* mRNA levels were at least 10 fold lower in GPI- B lymphocytes than in their GPI+ counterparts and in B lymphocytes from the heterozygous individual (2E). *PIGM* mRNA levels

were approximately 20 times lower in the homozygous GPI- T lymphocytes than in their correspondent GPI+ T lymphocytes and 40 times lower than in the heterozygous T lymphocytes. Noteworthy, *PIGM* mRNA expression in the GPI+ monocytes and in the GPI+ erythroid precursor cells was 4 and 5 times higher, respectively, in the homozygous individual (2C) than in the heterozygous (2E).
Curiously, heterozygous GPI+ monocytes expressed 2 times less *PIGM* mRNA than homozygous GPI- B lymphocytes.
Due to scarcity of primary samples these results derived from one extraction (2C monocytes, and CD36+ erythroid cells) or from two biological replicates (B lymphocytes and T lymphocytes).

Given the small number of primary cells from individuals of Family 1, *PIGM* expression was evaluated in lymphoblastoid B cell lines (LBCLs) derived from the same individuals. In parallel, measurements were also performed in LBCLs derived from individuals of Family 2 and normal individuals. In both families, *PIGM* levels in cells carrying the -270C>G mutation (1B and 2C) were less than 1% of those in normal homozygous (1A and 2D) and heterozygous (2E) B cells (Fig. 20A). *PIGM* mRNA levels correlated with GPI expression at the cell surface, as assessed by staining with FLAER (Fig. 20B). Moreover, in LBCLs from both families, expression of *PIGX* mRNA varied less than 2-fold with no significant changes between cells from the homozygous affected children (1B and 2C), homozygous WT (1A), heterozygous (2E and 2D) and normal controls (N1 and N2) (Fig. 21), suggesting that PIGX does not compensate for the deficiency in PIGM.

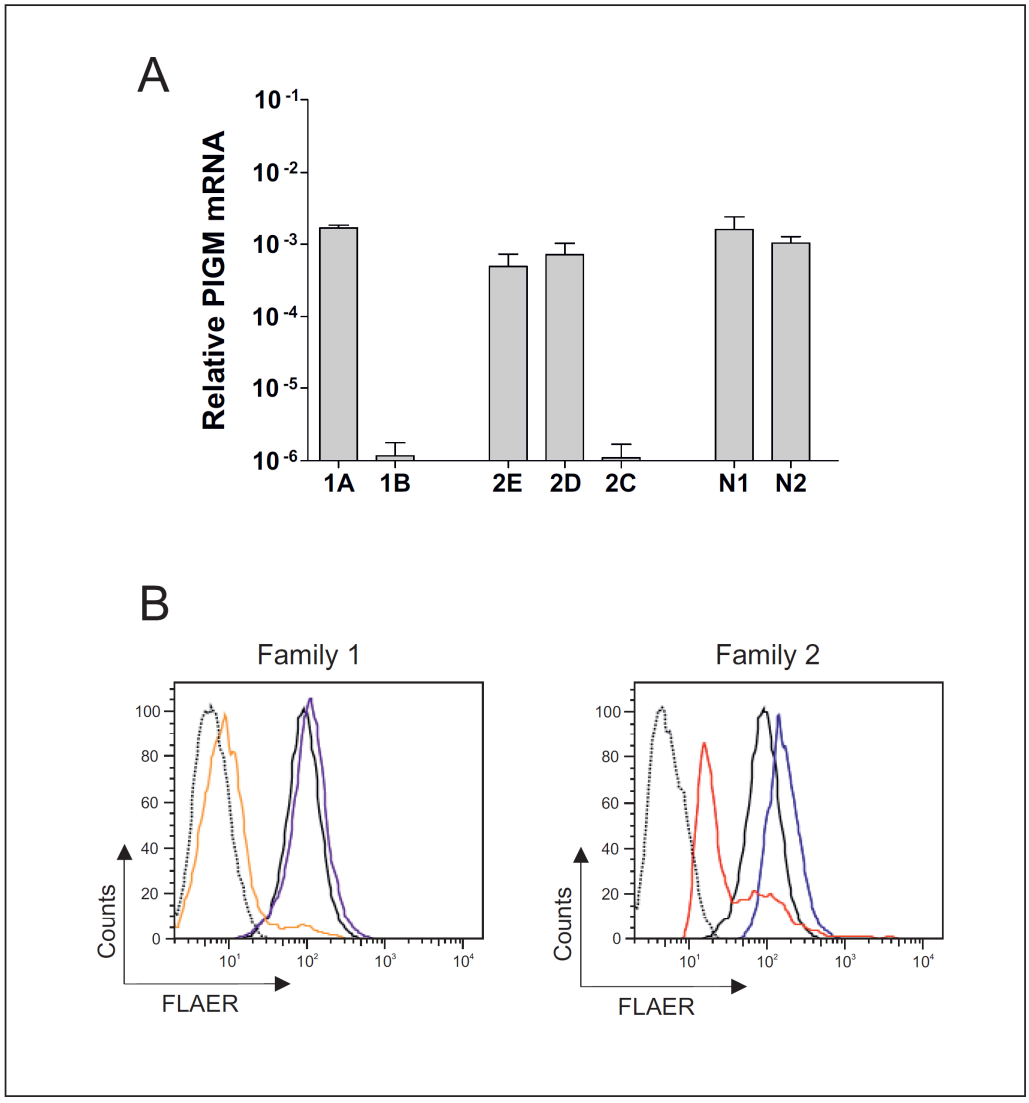


Fig. 20 – *PIGM* mRNA and GPI levels in EBV-transformed lymphoblastoid B cells (LBCLs) derived from individuals of Family 1 and Family 2. A - *PIGM* mRNA levels in LBCLs derived from the homozygous affected children (1B and 2C), homozygous normal individuals (1A and 2D), heterozygous individuals (2E) and normal controls (N1 and N2); see Fig. 9). Expression values were normalised to *GAPDH* and results are presented as mean ± SEM of 3 independent measurements, * $p<0.05$. **B -** GPI expression in LBCLs as assessed by staining with FLAER. Family 1: dotted line – unstained; solid orange line – homozygous MUT 1B; solid purple line – heterozygous 1A; solid black line – N2. Family 2: dotted line – unstained; solid red line – homozygous MUT 2C; solid blue line – heterozygous 2E; solid black line – N2.

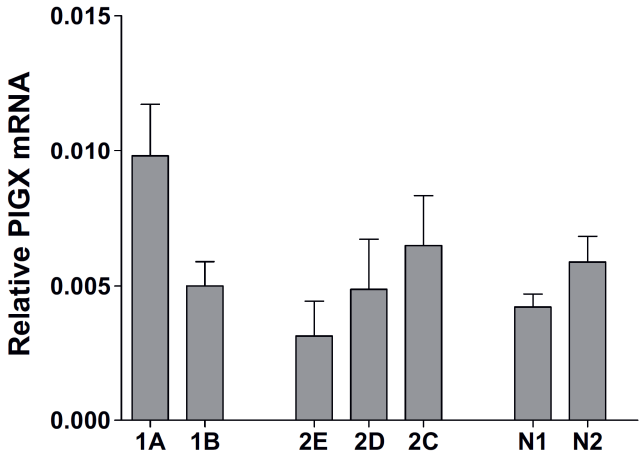


Fig. 21 – *PIGX* mRNA expression levels in LBCLs derived from the homozygous affected children (1B and 2C), homozygous normal individuals (1A and 2D), heterozygous individuals (2E) and normal controls (N1 and N2); see Fig. 9). Expression values were normalised to *GAPDH* and results are presented as mean ± SEM of 3 independent measurements.

Altogether these results show that in IGD haematopoietic cells and cell lines, the relationship between *PIGM* mRNA levels and efficient synthesis and expression of GPI is different between cell types.

My findings also suggest that in the erythroid cells and monocytes the -270GC rich motif is not essential for *PIGM* transcription.

To further investigate the impact of the pathogenic mutation in erythroid cells, I performed a transactivation assay. Almeida *et al*, 2006 (87) showed that in LBCLs and in the non-haematopoietic cell line HeLa, the activity of a 2Kb-*PIGM* promoter construct harbouring the -270C>G mutation (MUT) is reduced by 80-90% and 66%, respectively, in relation to normal WT promoter activity. I found, that whilst in HeLa cells the mutation caused a reduction to less than 50% of the WT promoter activity, in the K562 erythroid cells, the MUT promoter was nearly as active as the WT promoter (Fig. 22) thus further corroborating my observation in primary erythroid cells, i.e., that the pathogenic mutation does not decrease transcription of *PIGM* in erythroid cells.

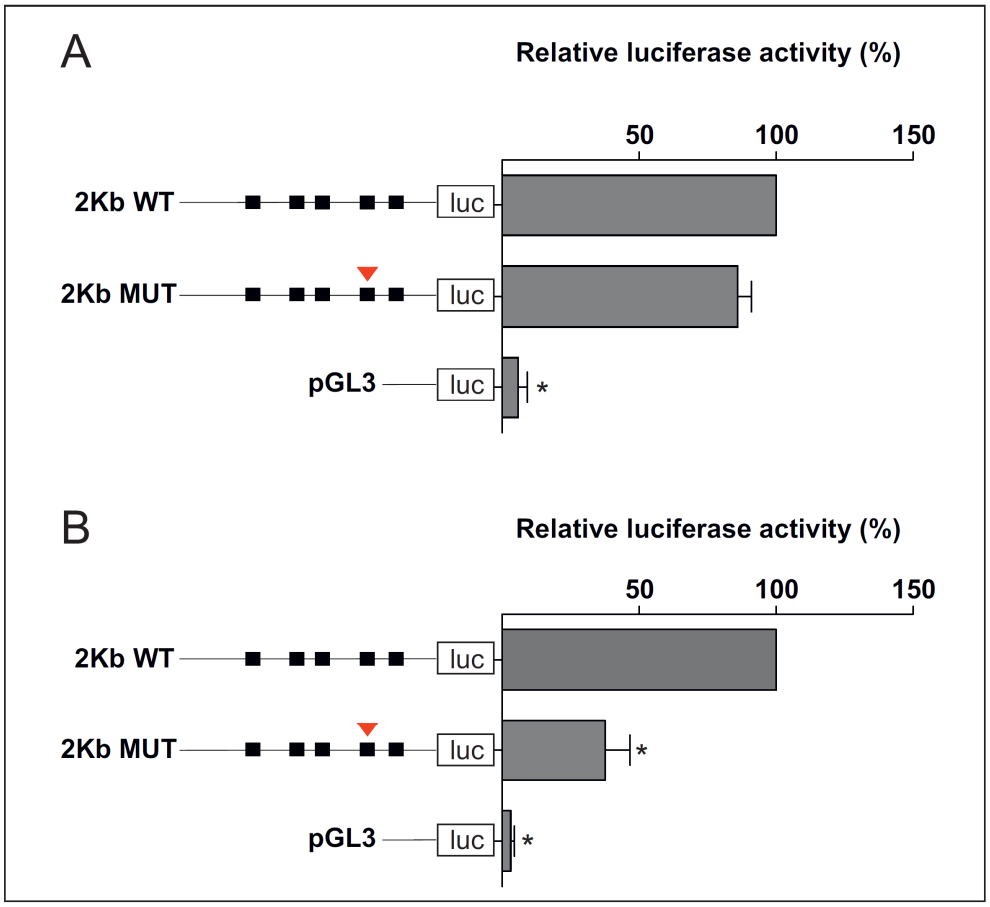


Fig. 22 – Luciferase reporter assay in K562 (A) and HeLa (B) cells. 48 hours after transfection cell lysate was assayed for luminescence for a period of 100ms. Luminescence values were normalised to renilla and results are presented as fold change relative to 2Kb-WT promoter ± SEM of 3 independent experiments measured in triplicate, * $p<0.05$.

The finding that the -270C>G mutation does not abrogate *PIGM* promoter activity (Fig. 22) and GPI expression in the erythroid cells (Fig. 18B), pointed towards potential mechanisms accounting for intact *PIGM* transcription in this cell type. Before proceeding with the exploration of different hypotheses such as the role of cell-type specific TFs in the erythroid-specific transcriptional regulation of *PIGM*, I studied the architecture of its promoter in erythroid cells.

PIGM promoter architecture influences differential expression of *PIGM*

Chromatin accessibility, that is determined by factors like DHSs and histone marks, is known to influence TF binding at the gene regulatory regions and therefore to impact on gene expression. Conversely, we recently showed that disruption of Sp1 binding influences the chromatin accessibility at the *PIGM* promoter. By employing a restriction enzyme accessibility assay (REAA) we showed that the MUT *PIGM* core promoter was highly protected and approximately 6-fold less accessible to digestion than the core promoter in normal B cells, associated with nucleosomal compaction and transcriptional repression. Treatment with NaBu, prompted nucleosome relaxation, histone re-acetylation and *PIGM* transcription by restoring Sp1 binding (38).

Following this observation, nucleosome occupancy in the core promoter of *PIGM* was evaluated in erythroid, epithelial and B cell lines. For that purpose, a micrococcal nuclease protection assay, adapted from (181) was employed. This method takes advantage of the micrococcal nuclease (MNase) enzyme that preferentially digests the DNA linker region of chromatin between individual nucleosomes, hereafter called mononucleosomes. Optimal MNase concentration was monitored in an agarose gel (see Appendix A, Fig. A3) and subsequently used to obtain the mononuclesome units. After DNA purification and by placing primers along 600bp of the core and proximal promoter of *PIGM*, nucleosome occupancy was determined by qRT-PCR in which 100bp amplicons overlapped in 50bp. High amplification PCR rates corresponded to protected, well positioned nucleosome regions whereas low amplification rates corresponded to more open and sensitive to digestion, less positioned regions.

The results showed that in the WT LBCL N2, nucleosome occupancy was higher than in K562 cells or epithelial HeLa cells, particularly, in the region encompassing the -270GC rich motif and the putative TSS. To ensure these findings were specific to *PIGM*, nucleosome occupancy at the core promoter of the housekeeping gene *GAPDH* was assayed in parallel. Nucleosome occupancy at the *GAPDH* promoter was uniformly low in all cell types tested (Fig. 23A).

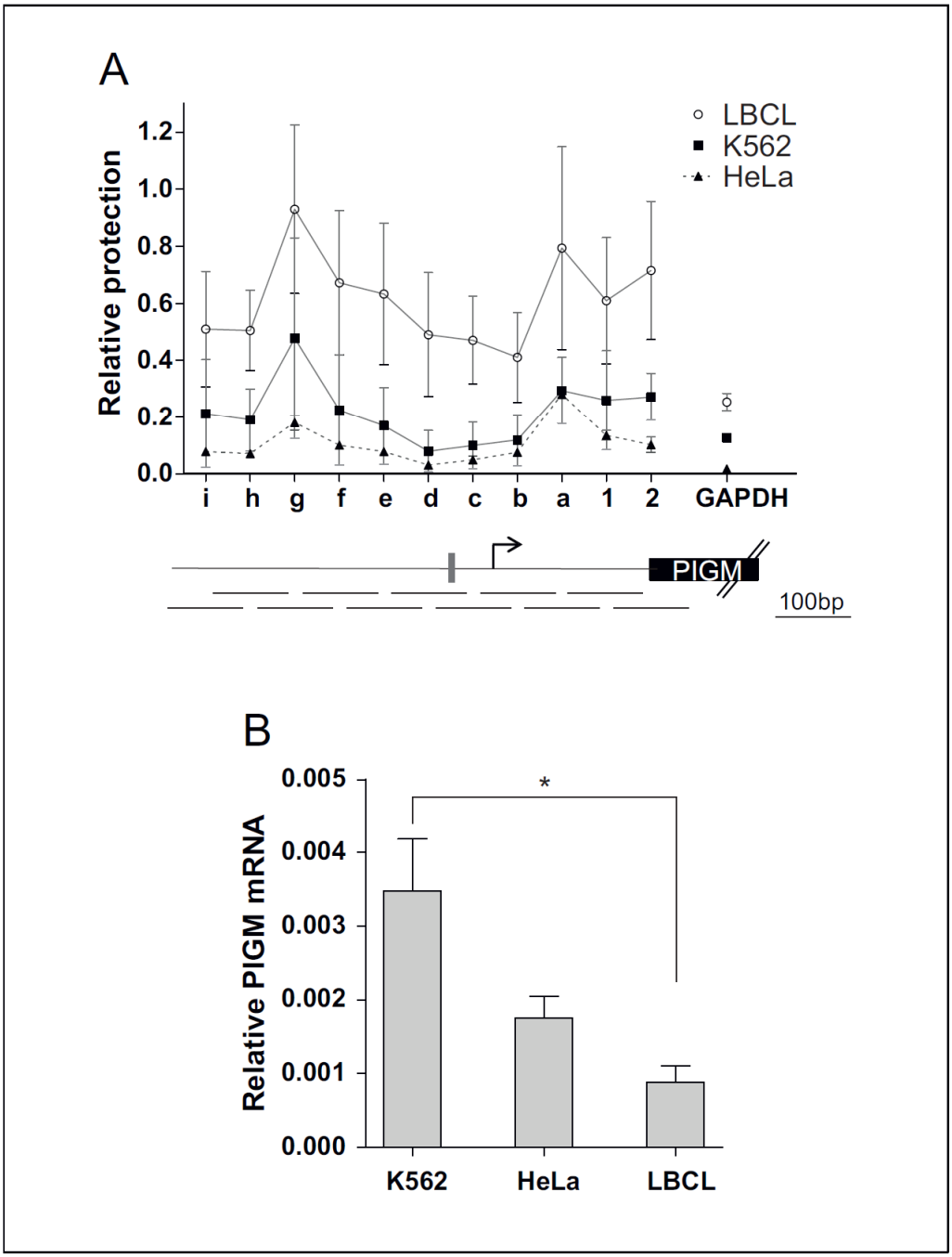


Fig. 23 – Effect of nucleosome occupancy on *PIGM* transcription. A – Nucleosome occupancy in the core and proximal promoter of *PIGM*. **i** to **2** represent the primers used for amplification, overlapping in 50bp. Nucleosome occupancy was determined by calculating the ratio of digested to undigested chromatin for equivalent DNA amounts. *GAPDH* promoter was used as negative control. **B** – *PIGM* mRNA expression levels in K562, HeLa and normal LBCL N2. Expression values were normalised to *GAPDH* and results are presented as mean ± SEM of 3 independent extractions, * $p<0.05$.

Since nucleosome protection at the *PIGM* promoter was higher in B cells than in erythroid or HeLa cells, its influence on *PIGM* transcription was evaluated by qRT-PCR analysis. Indeed, *PIGM* mRNA expression was lower in B cell lines than in K562 and HeLa cells (Fig. 23B), showing a clear link between nucleosome compaction and transcriptional output and suggesting that the differentially organised nucleosomes at the *PIGM* promoter impact on *PIGM* transcription in a cell type-specific manner.

To further dissect the architecture of the *PIGM* promoter, an attempt was made to determine with precision the TSS. In principle the presence of a TSS in erythroid cells upstream of the -270GC motif can potentially 'bypass' the pathogenic mutation and ensure intact *PIGM* transcription in these cells.

In the past years, several methods have been used to capture capped mRNAs and map the 20-30 nucleotides of the complementary DNA, thus defining TSSs (reviewed in (21)). One of these methods is 5'-Rapid Amplification of cDNA Ends (5'-RACE). Briefly, after isolation of capped mRNA, its 5'-end is tagged and reverse-transcribed into cDNA following annealing with an oligo(dT) sequence. The second cDNA strand is then synthesised and the double-stranded cDNA (dscDNA) is amplified in a PCR reaction by using specific primers. As a control, dscDNA is tested for detection of the abundant β-actin mRNA and in parallel a PCR reaction is carried out by using primers specific for the gene of interest. The PCR product is then cloned into a vector backbone and sequenced by using specific primers for the gene of interest (see Chapter 2, 2.13).

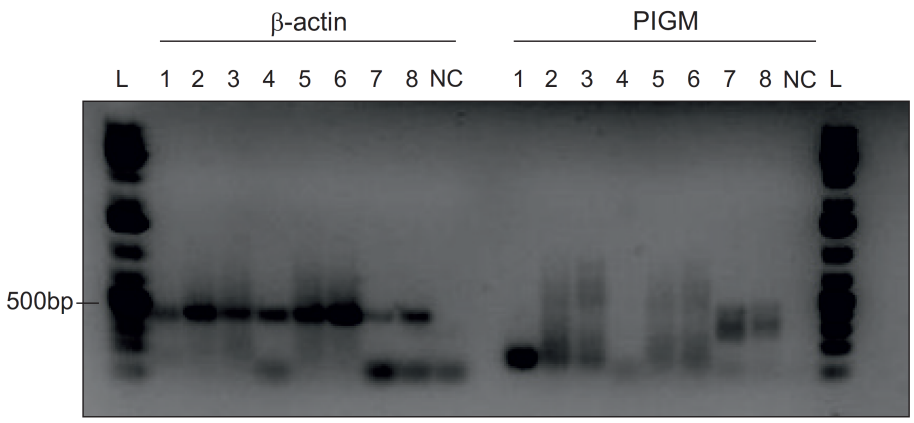


Fig. 24 – Gel electrophoresis of the 5'-RACE-PCR. Electrophoresis in a 2% agarose gel of β-actin dscDNA PCR (left) and PIGM dscDNA PCR (right). 1 – WT LBCL (N2), 2 – Jurkat, 3 – HL60, 4- Fibroblast, 5 – K562, 6 – SH-SY5Y, 7 – HeLa, 8 – HeLa (commercial), NC – negative control, L - ladder. The β-actin positive control PCR generates an amplicon of approximately 400bp.

The size of the PCR amplicon derived from the amplification with the β-actin primers was consistent with the predicted size (approximately 400bp). However, in the PCR reaction using primers targeting *PIGM*, the result in the majority of the cell lines was not clear, suggesting unspecific amplification or the presence of more than one TSS (Fig. 24). After several attempts to optimize the PCR conditions (by altering the annealing temperature and also by using DMSO, as it disrupts base pairing thus facilitating strand separation), the results obtained were similar (data not shown). Following cloning of the PCR product from K562, WT LBCL and HeLa into the pJET vector and transformation of competent cells, the plasmid DNA was sequenced. The results showed a very small number of sequences mapping to *PIGM*. While in the erythroid cell line K562, two of the one hundred sequences aligned with *PIGM*, in the WT B cell line none of the sequences obtained from the one hundred colonies aligned with the gene. In HeLa cells the 5'-RACE reaction was more efficient as approximately 30% of the 20 sequences analysed matched with *PIGM* gene and some inclusively matched between them (Fig. 25).

Although inefficient, these results showed that in the cells analysed so far, the TSSs are positioned downstream at the -270GC motif, excluding the possibility of transcription initiation upstream of the pathogenic mutation.

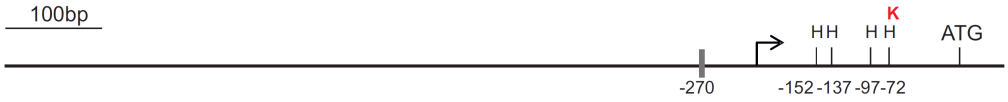


Fig. 25 – Schematic representation of the TSSs mapped at the *PIGM* promoter in K562 (K) and HeLa (H) cells. In K562 the TSS localised -72bp upstream of the ATG and in HeLa four TSSs were identified with the following positions, -72, -97, -137 and -152bp upstream of the ATG. Black arrow – putative TSS (87). Grey bar – -270 Sp1 binding site.

Furthermore, CAGE (Cap Analysis Gene Expression) data deposited in the online database ENCODE revealed the presence of two clusters of TSSs at the promoter in both the B cell line GM12878 and the erythroid K562 cells. Each cluster consists of several TSSs, suggesting that transcription initiation might take place at different positions. Nevertheless, both clusters localised downstream at the -270GC motif, and were similar between the erythroid and B cells (Fig. 26).

Taken together, these results show that all TSS are downstream of the -270 motif position and thus cannot account for intact *PIGM* transcription in erythroid cells.

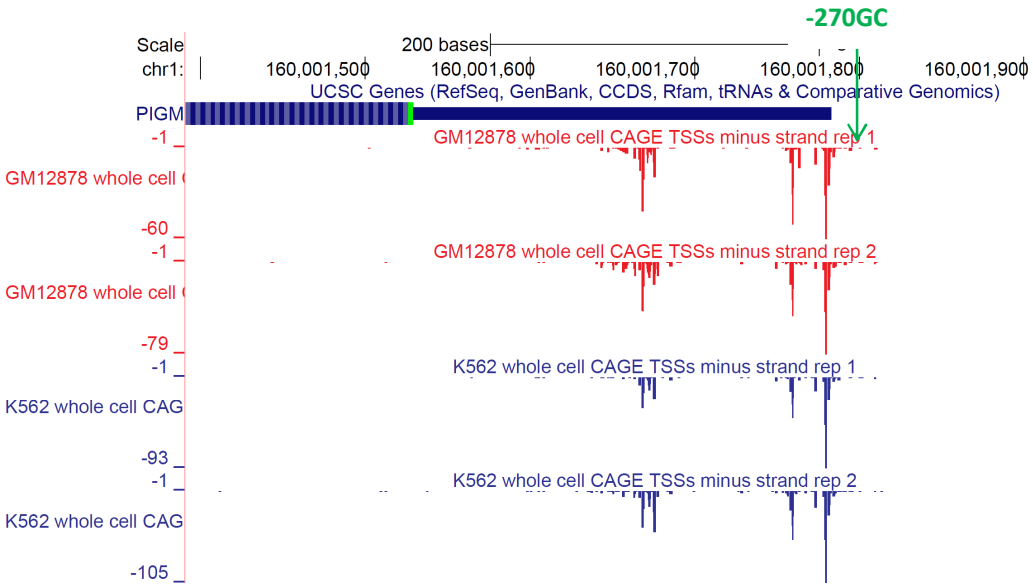


Fig. 26 – CAGE analysis of *PIGM* (strand -1) in the B cell line GM12878 (red) and in K562 (blue) retrieved from the ENCODE database. The TSSs are grouped in clusters in which the height of the peaks is proportional to the number of matched sequences. The coordinates for the -270GC motif are chr1:160,001,799 – 160,001,804 (green arrow). Both groups of TSSs are localised downstream of the -270GC motif. The method allows for single nucleotide resolution (21).

3.1.4 Discussion

GPI moieties are structures indispensable for anchoring of proteins at the cell surface which in turn, are required for the maintenance of several functions of the cell. PIGM constitutes the catalytic component of the GPI MT-I (68) whereas PIGX is believed to be responsible for maintaining the stability of the complex (69). The GPI MT-I is essential for the GPI biosynthesis as repression of *PIGM* transcription abrogates GPI expression.

Since *PIGM* is a housekeeping gene, it is constitutively expressed in all cell types and it is fundamental for the basic functions of the cell (182). The concept of housekeeping gene has been recently revisited by Eisenberg *et al*, 2013 (183), in which the authors emphasise that a housekeeping gene has to be expressed at a constant level across cell types. Here, I challenge the concept proposed by the

authors as I observed that expression of *PIGM* varies considerably between different primary cells and between normal individuals (Fig. 15).
Very interestingly, my results show that haematopoietic cells have different basal mRNA levels. For instance, whereas *PIGM* levels in the monocytes of the children with IGD were sufficient for normal GPI expression, in the B lymphocytes those levels were not sufficient and GPI was not produced (Fig. 19). Assessment of PIGM protein levels would have been useful to understand the basis of these differences.

In erythroid lineage cells, *PIGM* mRNA levels decreased as normal erythroid precursor cells underwent differentiation *in vitro* (Fig. 15). Although the result seems plausible due to the fact that gene expression is "shut-down" as erythroid cells proceed towards late stages in the erythropoiesis, it has to be taken into account that the analysis was not performed in well-defined, pure, CD36+GlyA+ populations but in whole $CD36^{+}$ erythroid precursor cells that were heterogeneous in terms of CD71 and GlyA expression. Nevertheless, in primary murine fetal erythroid precursor cells sorted at different stages of differentiation according to Ter119 and CD71 expression, the same result was described. As seen by RNA-seq, *PIGM* expression decreased as erythroid cells become more mature (Fig. 16).

However, erythroid precursor cells defined as CD36+GlyA+, homozygous or heterozygous for the pathogenic -270C>G mutation expressed normal levels of the GPI-linked antigen CD59 (Fig. 18B), clearly demonstrating that in this well-defined erythroid precursor population the level of *PIGM* mRNA even in homozygous mutant cells is not only comparable (or in fact somewhat higher) than in normal cells (Fig. 19) but it is also sufficient for normal GPI expression. The finding that the homozygous or heterozygous erythroid precursor cells express slightly more CD59 than the normal cells is most likely related with normal variation between healthy donors.

Nevertheless, there are two drawbacks in relation to these results: firstly, due to scarcity of primary samples I could only evaluate one patient sample and secondly, PIGM protein level assessment was not possible due to lack of suitable

antibody. Therefore, I could not provide direct evidence that variable *PIGM* mRNA expression translates into variable PIGM protein expression.

Mutations in gene regulatory regions have been described in the literature and very frequently generate or abolish TF binding sites (TFBS).
We have recently showed that in the B cells, the -270C>G mutation represses *PIGM* transcription by abolishing Sp1 binding and facilitating Polycomb recruitment. Here, I showed that in normal WT B cells, high levels of nucleosome protection and thus nucleosome occupancy in the region encompassing the -270GC motif correlated with low levels of *PIGM* transcription, reinforcing the link between chromatin accessibility and transcriptional output. In contrast, in the erythroid cells the -270 GC-rich motif was not required for *PIGM* promoter activation, a finding supported by the luciferase reporter assays (Fig. 22) and the fact that the promoter region encompassing the -270GC motif was highly accessible for transcription (Fig. 23A and B). Further investigations, employing for instance a DNase footprinting assay, would have helped to reinforce the finding that the -270GC motif is transcriptionally accessible. Given the differences in chromatin status and transcriptional potential, I propose that the presence of the pathogenic C>G mutation is sufficient to tip the promoter into a repressive state in B but not erythroid cells, thus in part explaining the differential cellular phenotype in IGD.

PIGM falls in the category of the genes with GC-rich promoters, which are commonly associated with multiple TSSs over a region of 50-100 nucleotides. By using ‘5-RACE, I have shown that in K562 cells the TSS localised -72bp upstream of the ATG and in HeLa four TSSs were identified (Fig. 25). Given the high percentage of false positives, possibly due to the inefficient PCR reaction, robust conclusions could not be drawn. However, by using CAGE data, two clusters of TSS upstream at the -270C>G mutation were annotated in both B and erythroid cells (Fig.26). Whether they correspond to alternative promoters is not known but so far the data revealed that the TSS of *PIGM* is positioned downstream of the pathogenic mutation. Therefore, I conclude that an alternative promoter operating upstream of the -270GC rich box cannot account for the intact transcription of *PIGM* in normal or patient erythroid cells.

Taken together, my results showed that in the erythroid cells, *PIGM* transcription is independent of the -270-Sp1 binding site. Understanding whether generic or erythroid-specific TFs can overcome the -270C>G mutation is the focus of the next chapters.

4.1 Exploring the role of the erythroid-lineage affiliated TFs GATA-1 and KLF1 in *PIGM*-associated IGD

4.1.1 Introduction

In *PIGM*-associated IGD, red blood cells express near normal expression of GPI, hence the lack of haemolysis and anaemia in patients (87). In the previous chapter I showed that in erythroid cells the -270GC-rich motif is not essential for *PIGM* promoter activity and transcription, hence the normal expression of GPI. The results suggested that in erythroid cells *PIGM* transcription is regulated in a cell-type specific manner.

cis-DNA elements control transcriptional regulation through the action of TFs bound to these elements. The role of the erythroid-restricted TFs GATA-1 and KLF1 as master regulators of erythroid transcriptional programmes (145, 175) and their implications in disease (170, 174, 184) has been documented in mice and in humans. Interestingly, *in vitro* experiments have suggested that regulation of erythroid specific genes by GATA-1 often occurs in co-operation with Sp1 (134, 185) and in addition, GATA-1 and Sp1 can physically interact through their zinc-finger DNA binding domains (135, 168). Moreover, *in vitro* assays have shown that according to the erythroid promoter context, GATA-1 can interact either with Sp1 or KLF1 (165). KLF1, a TF predominantly expressed in erythroid cells, shares the same GC-rich binding sequence with the Sp family of TFs, with a preference for the 5'-CACCC-3' DNA motif (186).

The aim of this chapter is **to investigate the mechanism(s) responsible for intact *PIGM* transcription and normal GPI expression in erythroid cells.** Here, I hypothesised that the erythroid lineage-specific TFs GATA-1 and KLF1 drive *PIGM* activity in IGD erythroid cells even in the presence of the pathogenic -270C>G mutation.

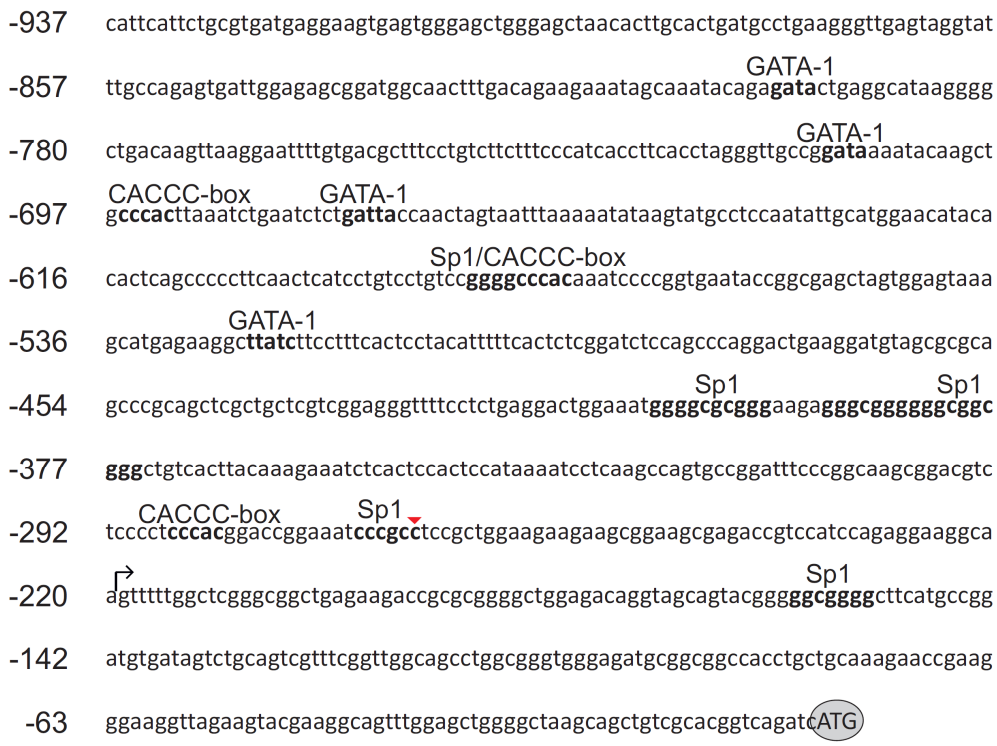
4.1.2 Experimental design

To test the above hypothesis, the *PIGM* promoter was subjected to a bioinformatic analysis for putative TFs binding sites. The role of *GATA-1* on *PIGM* transcriptional regulation was then assayed by ChIP analysis in the erythroid cell line K562 and in primary erythroid precursor cells followed by a genetic approach in which I employed knockdown and overexpression assays. KLF1 occupancy at the *PIGM* promoter was assayed by ChIP-qPCR in primary erythroid precursor cells.

4.1.3 Results

GATA-1 is not a transcriptional activator of *PIGM*

By employing a bioinformatic analysis using the Transcription Element Search System (TESS) and the TFsearch (http://www.cbrc.jp/research/db/TFSEARCH.html) web tools, four putative GATA-1 binding sites were predicted at the -799bp, -712/-675bp and -520bp (from ATG) at the *PIGM* promoter (Fig. 27). Although the bioinformatics analysis has not predicted the presence of putative KLF1 binding sites, several CACCC-DNA motifs were identified upstream at the -270C>G mutation (-696bp, -578 and -291). The presence of four putative Sp1 binding sites (-160, -395/-380, -582) in addition to the validated -270 Sp1 motif has been already described in Almeida *et al*, 2006 (87).



Predicted sites	TESS	TF Search (out of 100)
GATA-1 (-799)	12	88.1
GATA-1 (-712/-675)	16	94.3
GATA-1 (-523)	10	90.4

Fig. 27– Bioinformatic analysis of the *PIGM* promoter. Analysis was performed in the promoter region encompassing -2000 to +1 (ATG) using the bioinformatic tools TESS and TFsearch. **A** - *PIGM* promoter sequence showing the TFBS predicted form the analysis. The presence of four GATA-1 sites (bold) were predicted at the following positions -799, -712/-675 and -523. In addition to the -270 Sp1 motif, four putative Sp1 sites, already described in Almeida, A. *et al*, 2006, were predicted (-582, -395/-380 and -160). Three CACCC boxes were identified according to their sequence. ATG – translation initiation. ▼-270 C>G mutation. Arrow – putative TSS. **B** – Table with GATA-1 binding scores. Cut-off score TFsearch = 85.0.

High throughput ChIP-seq data deposited in the ENCODE database revealed GATA-1 binding at the *PIGM* promoter in the cell line K562 and in primary

erythroblasts derived from PBMC (Fig. 28), thus reinforcing the bioinformatics analysis.

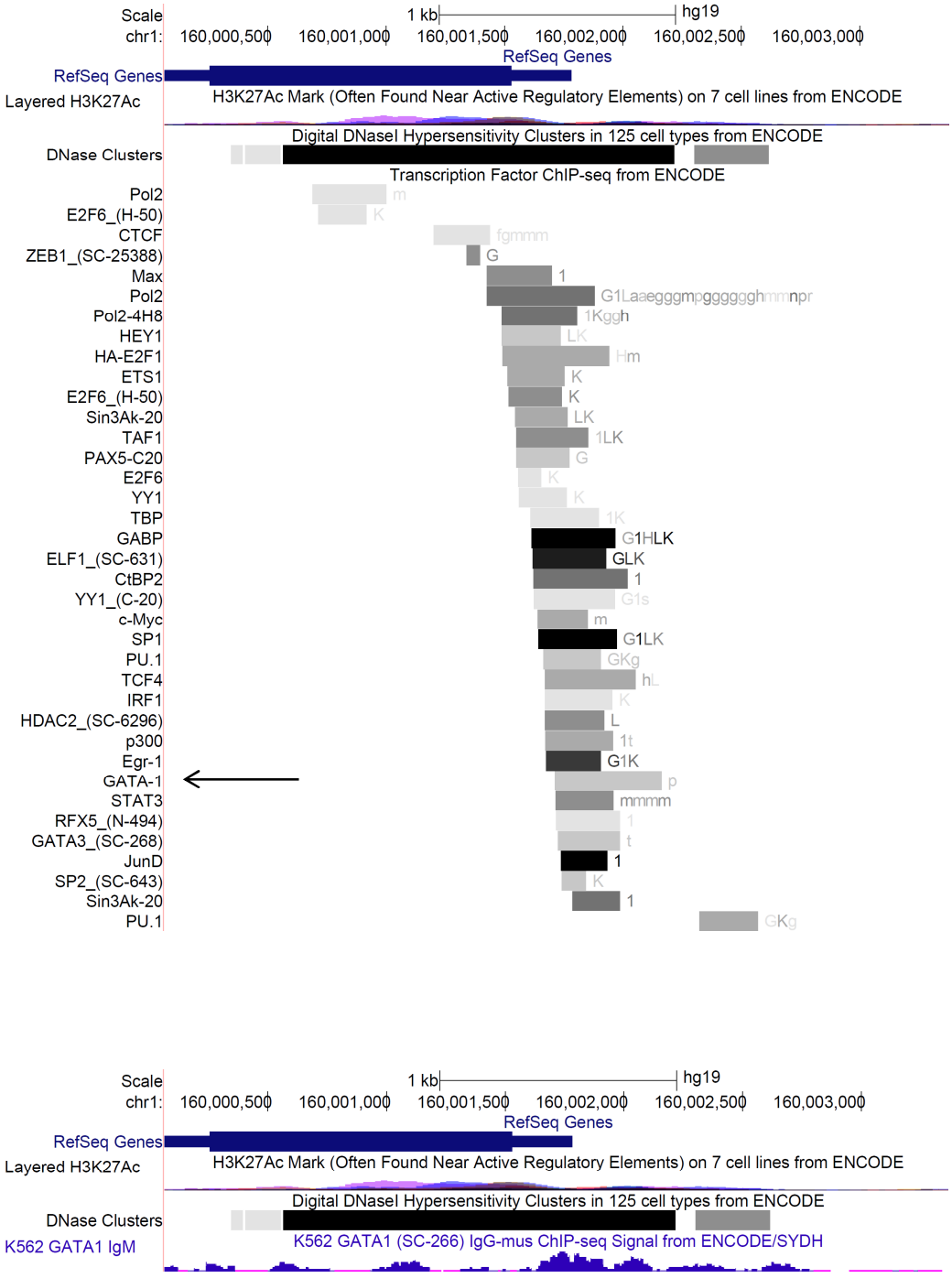


Fig. 28 – ChIP-seq analysis at the *PIGM* promoter retrieved from the UCSC genome browser on Human Feb.2009 (GRCh37/hg19) Assembly - ENCODE database. GATA-1

binding was captured in primary erythroblast precursor cells derived from PBMC (top). GATA-1 binding at the *PIGM* promoter in K562 (bottom).

Considering the great interest in dissecting GATA-1 occupancy at regions set apart by less than 200bp at the *PIGM* promoter, ChIP analysis followed by qPCR was carried out in which site-specific primers were designed to flank the predicted GATA-1 sites (Fig. 29). For that purpose, the well-studied erythroid cell line K562 was selected. GATA-1 was highly expressed in K562, whereas in cell lines with epithelial, myeloid, lymphoid or neuronal origin GATA-1 was not expressed (Fig. 30).

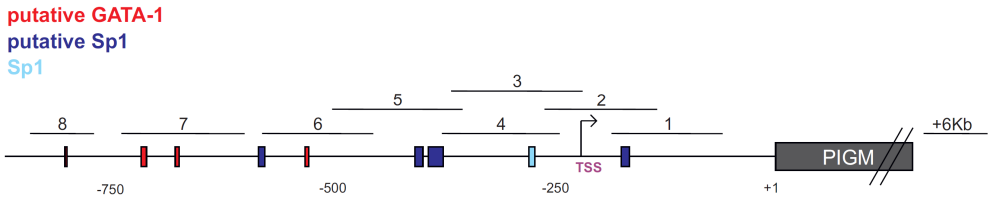


Fig. 29 – Schematic representation of the *PIGM* promoter. Red – putative GATA-1 binding sites; blue – putative Sp1 binding sites; light blue – Validated Sp1 binding site at the -270GC box. 1 to 8 and +6Kb designate the PCR amplicons. Two sets of primers were used to generate two amplicons (3 and 4) spanning the -270GC box.

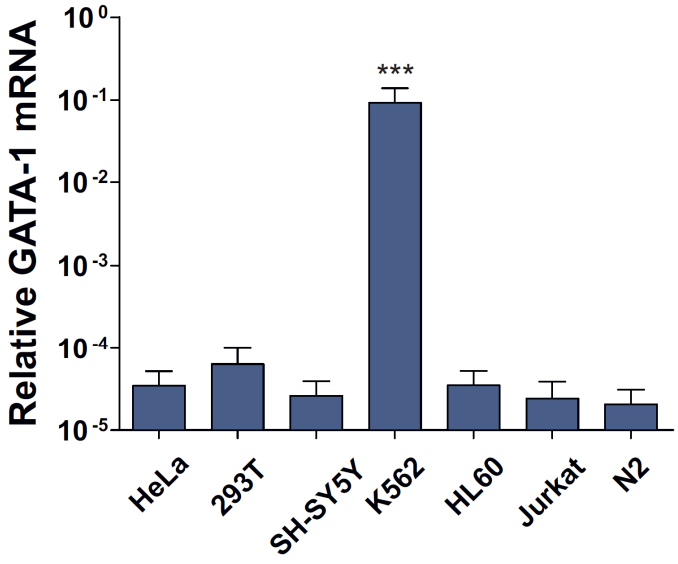


Fig. 30 – *GATA-1* mRNA expression levels in cell lines from epithelial (HeLa, 293T), neuronal (SH-SY5Y), erythroid (K562), myeloid (HL60), T lymphocytic (Jurkat) and B lymphocytic (LBCL N2) origin. Expression values were normalised to *GAPDH* and results are presented as mean ± SEM of 3 independent extractions, *** $p<0.001$.

Typically, chromatin is prepared and sonicated (see Chapter 2, 2.5) in fragments between 200bp-600bp (187) that are visualised on agarose gel. Following sonication (Appendix A, Fig. A4), chromatin immunoprecipitation was carried out with 5µg of GATA-1 antibody or respective IgG control and the promoter was scanned for GATA-1 binding. GATA-1 enrichment was compared to GATA-1 binding at its own gene regulatory areas. *GATA-1* is regulated by DNase I Hypersensitive sites (HSs) as described in Valverde-Garduno *et al*, 2004 (188). Whereas HS -3.5 and HS +14 are proximal to GATA-1, HS +15 is located more distally at the 5'end of the *HDAC6* gene. Accordingly, I found that in K562 cells, GATA-1was enriched 36 and 27 fold at the HS -3.5 and HS +14, respectively, and less enriched at the HS-49 and HS+15.

At the *PIGM* promoter, GATA-1 was moderately (7 fold) enriched at the predicted GATA-1 sites (amplicons 8, 7 and 6) when compared to the gene desert region, GW10.

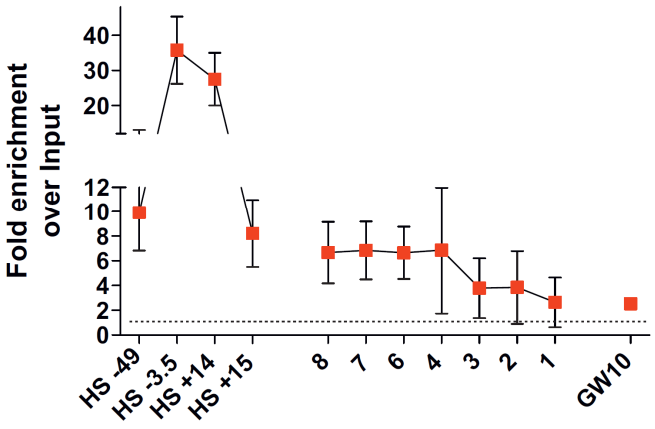


Fig. 31 – GATA-1 occupancy at the length of the *PIGM* promoter in K562. HS -49, HS -3.5, HS +14 and HS +15 represented the DNase hypersensitive sites at *GATA-1* locus used as controls. GW10 – negative control. Dashed line – IgG control. IgG and GATA-1 enrichment were expressed as % of Input. Results are presented as fold enrichment over input IgG ± SEM of 5 independent experiments.

To further validate this finding, the same analysis was performed in primary $CD36^{+}$ erythroid precursor cells differentiated *in vitro* from PBMC.

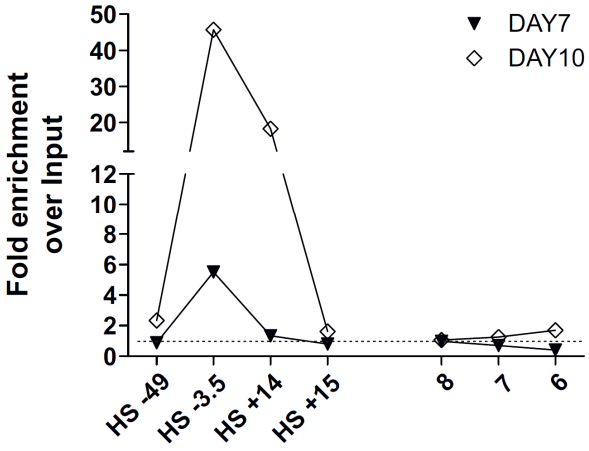


Fig. 32 – GATA-1 occupancy at the length of the *PIGM* promoter in primary erythroid precursor cells. Cells were differentiated *in vitro* from PBMC as shown in Fig. 14. HS -49, HS -3.5, HS +14 and HS +15 represented the DNAse Hypersensitive

sites at *GATA-1* locus used as controls. Dashed line – IgG control. IgG and GATA-1 enrichment were expressed as % of Input. Results are presented as fold enrichment over input IgG n=1.

Erythroid differentiation was monitored by flow cytometry analysis as described before (see Results 3.1). Briefly, cells were analysed for CD34, CD36, CD71 and GlyA expression and isolated by magnetic separation after staining with CD36-PE antibody and incubation with anti-PE beads. Morphologic characteristics of selected cells were assessed by MGG staining. After selection of the $CD36^+$ erythroid precursor cells, the remaining $CD36^-$ cells were used as a source of chromatin for optimisation of the appropriate sonication conditions (Appendix A, Fig. A4B). Following chromatin preparation of cross-linked $CD36^+$ cells and immunoprecipitation with anti-GATA-1 antibody and respective IgG control, qPCR analysis was performed. In erythroid precursor cells isolated on Day 7, GATA-1 was exclusively enriched at the *GATA-1* hypersensitivity site, HS-3.5 (6 fold) and later in differentiation, at Day 10, it was also enriched at the HS+14 (38 fold), suggesting a dynamic, time-coordinated, process. Remarkably, GATA-1 was not enriched at the four predicted GATA-1 sites (amplicons 8, 7 and 6) at the *PIGM* promoter (2 fold enrichment) (Fig. 32).

Regardless the differences seen between the erythroid cell line K562 and the erythroid precursor cells, the data suggested that the overall GATA-1 binding at the *PIGM* promoter was low.

In order to dissect whether GATA-1 has indeed a functional role in the transcriptional regulation of *PIGM*, at least in K562 cells, a genetic approach was employed. To this end, GATA-1 knock-down (kd) K562 cells were generated by using short hairpin RNA (shRNA) lentiviral particles. This method of delivery allows stable integration of the shRNA and therefore long-term silencing of the gene of interest. shRNAs consist of a 19-22bp sense sequence identical to the target mRNA, followed by a short loop of 4-11 nucleotides, and an anti-sense 19-22bp sequence. Following PolIII-mediated transcription, the shRNA is translocated to the cytosol where it is recognised by an enzyme called Dicer. The shRNA is then processed into a small-interfering RNA (siRNA) and once bound to the target RNA is incorporated into the RNA-induced silencing complex (RISC). Subsequently, the target RNA is degraded (189). In keeping with this,

GATA-1 mRNA target sequences were designed alongside with a scramble control by using a web tool (Mission®shRNA, Sigma – http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/shrna/shrna-search-and-order.html; see Chapter 2, 2.12). The synthesised oligonucleotide sequences were then annealed and cloned into the pLKO.1-GFP vector backbone (Appendix A, Fig. A5), which contains an ampicillin resistant marker for colony selection and a GFP reporter gene for selection of the transduced cells. Plasmid DNA was then digested to confirm the insertion of the shRNA (Appendix A, Fig. A6) and used to transfect, alongside with the helper viral vectors, 293T packaging cells that are able to produce the viral supernatant. Prior to transduction, the concentrated viral supernatant collected at 48h and 72h post-transfection was used for viral titration (Appendix A, Fig. A7). K562 cells were transduced with $4.5x10^5$ units/ml of virus and 96 hours post-transduction, GFP cells were sorted and cultured in the appropriate medium to recover.

Very interestingly, GATA-1 kd K562 cells lost their characteristic "red" colour as a result, most likely, due to down-regulation of Haemoglobin (Hb) genes. After expansion of the GFP cells, GATA-1 knockdown was assessed by measuring the mRNA and protein levels. As shown in Fig. 33, a reduction of 65% at the mRNA level in K562 GATA-1 kd cells resulted in a 20% down-regulation at the protein level, when compared with K562 scramble cells. As positive control for the effect of down-regulation of GATA-1 in K562, I used the γ globin gene (190). As shown, in K562 GATA-1 kd cells γ globin gene expression (γ exon2 and γ exon3) was reduced to less than 30% of the normal levels.

Strikingly, in stable K562 GATA-1 kd cells, *PIGM* mRNA levels were 2 fold higher than in scramble control cells, indicating that GATA-1 might be a repressor rather than an activator of *PIGM* transcription. Whereas Sp1 levels were comparable between K562 scramble and K562 GATA-1 kd cells, *PIGX* mRNA levels also increased 2-fold, consistent with a co-ordinate effect of GATA-1 on the transcriptional regulation of both genes involved in the first mannosylation of the GPI core structure (Fig. 34). An attempt to generate a second GATA-1 kd K562 cell line was made but it was unsuccessful due to the low knockdown efficiency (data not shown). Therefore, whether GATA-1 kd results indeed in upregulation of *PIGM* and *PIGX* mRNA requires further confirmation.

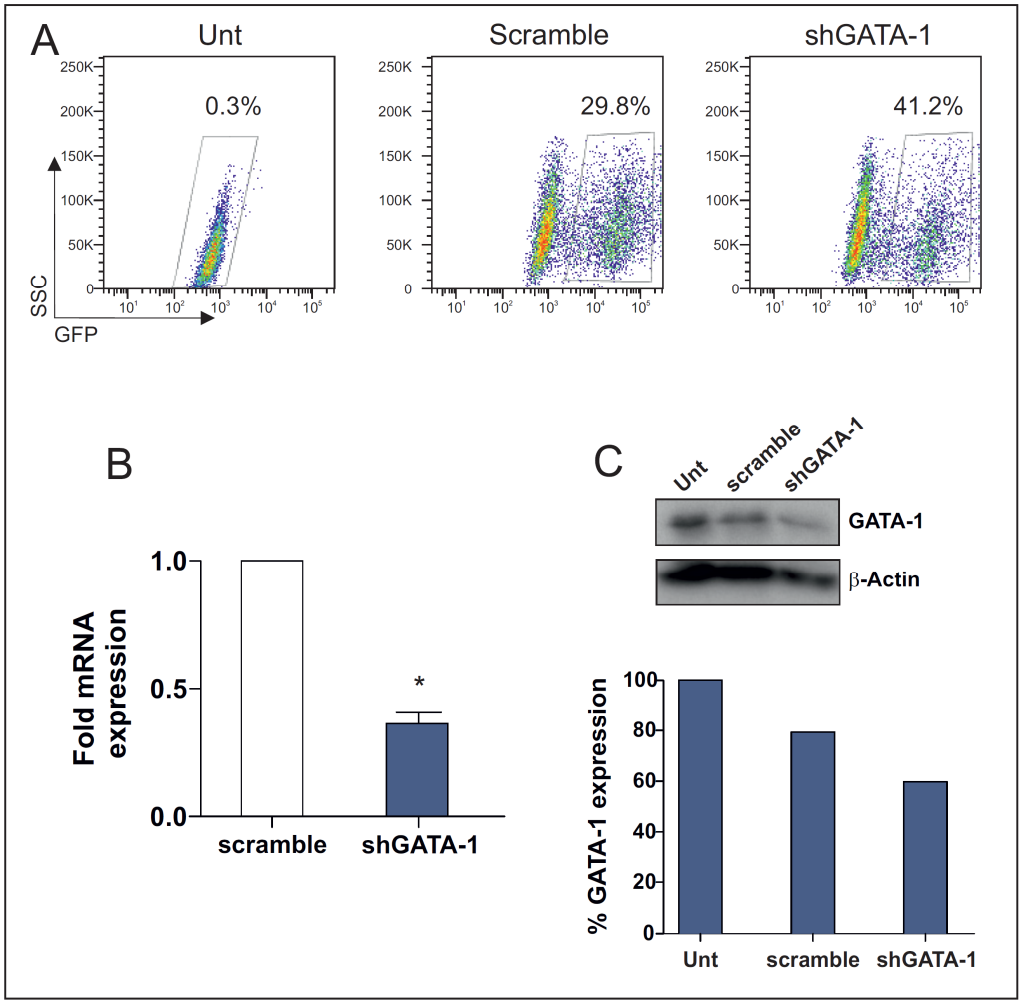


Fig. 33 – K562 cells transduced with short hairpin (sh) RNA targeting GATA-1 and control scramble. A – FACS plots show the efficiency of transduction as measured by GFP. Untransduced cells were used as unstained control sample. **B** – *GATA-1* mRNA expression in stable K562 scramble and K562 GATA-1 knock-down cells (K562 shGATA-1). Expression values were normalised to *GAPDH*. Results are presented as fold mRNA ± SEM of 3 independent extractions, ∗ $p<0.05$. **C** – GATA-1 and β-actin protein expression. % of GATA-1 expression was determined by using ImageJ software.

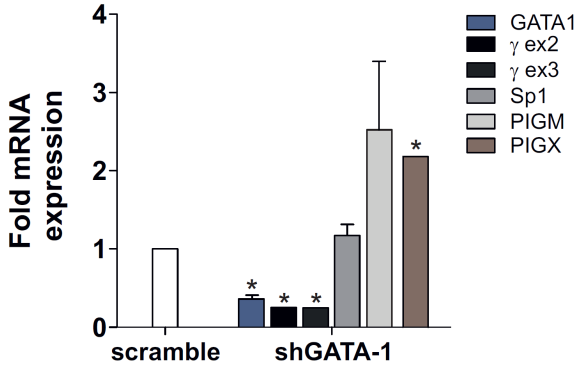


Fig. 34 – mRNA expression in K562 scramble and K562 GATA-1 knock-down cells. Expression values were normalised to *GAPDH*. Results are presented as fold mRNA ± SEM of three independent measurements,∗ $p<0.05$.

In an effort to better understand the role of GATA-1 in *PIGM* transcriptional regulation, a complementary study in which GATA-1 is overexpressed, was carried out. For this purpose, the Flp/FRT site-specific recombination system was used. By employing this method, a single copy of the -270GC WT or MUT *PIGM* promoter, driving a GFP reporter gene, could be integrated in a site-specific manner into the genome of the cells. The Flp/FRT method takes advantage of the Flp recombinase (pOG44) that mediates recombination of two FRT (Flp recombination target) sites. The FRT sites, one integrated in the genome of the Flp-In cell line and the other present in the pCDNA5/FRT vector (see Chapter 2, 2.9), constitute the binding and cleavage sites for the Flp recombinase.

Firstly, a sequence of 910bp (from ATG) of the *PIGM* promoter, either WT or MUT at the -270GC-rich motif, was sub-cloned upstream of a GFP reporter gene in the modified pCDNA5/FRT vector backbone (see Appendix A, Fig A1). The pCDNA5/FRT vector, now containing the *PIGM* promoter-GFP cassette was co-transfected with the pOG44 plasmid into the 293 Flp-In cells. Three weeks after transfection, recombinant, hygromycin resistant cells were obtained and expanded in the appropriate culture conditions. Thereafter, gDNA from both parental and recombinant cells was extracted and tested by using 3 primers-PCR (Fig. 35). Following confirmation of the recombination by PCR, the recombinant cell clones

were assessed for GFP expression by FACS but surprisingly, expression of GFP in those cells was very low and comparable to background levels (Fig. 36).

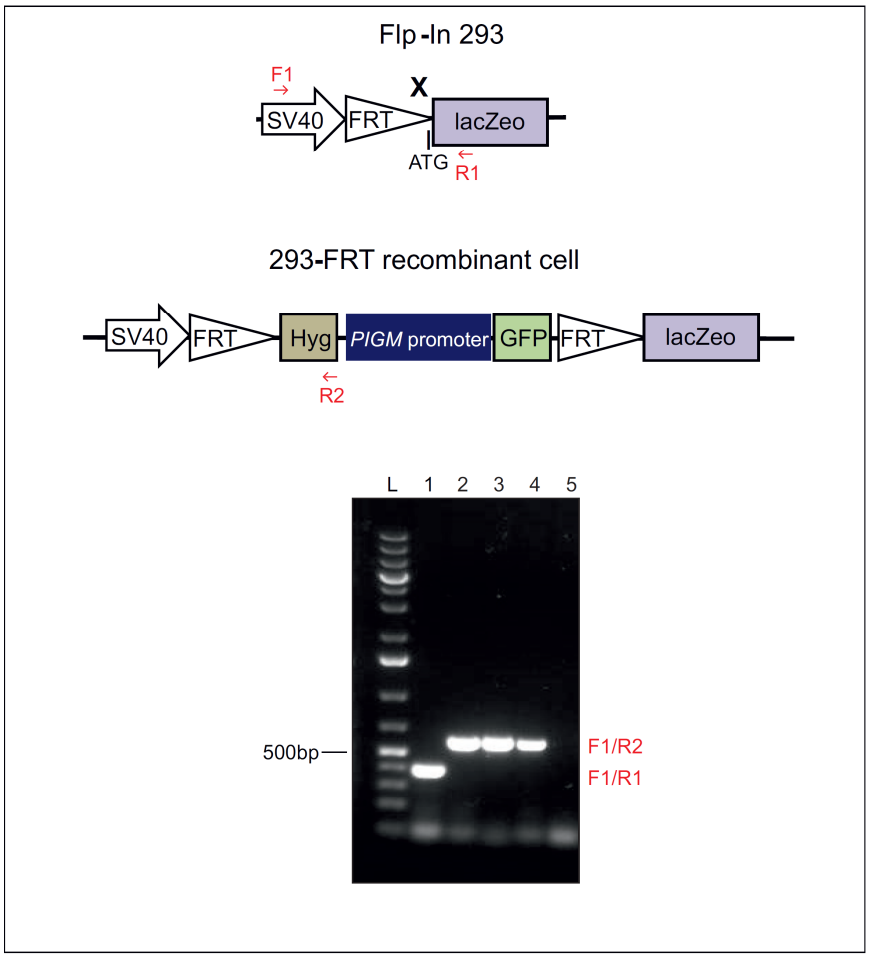


Fig. 35 – Gel electrophoresis of the PCR products of parental and recombinant 293-FRT amplified genomic DNA. L – ladder, 1 –Flp-In 293, 2 – 910WT *PIGM*-GFP cells, 3 - 910MUT *PIGM*-GFP cells (clone #1), 4 - 910MUT *PIGM*-GFP cells (clone #2), 5 – non-template PCR control. Primers placed on the *SV40*, *lacZeo* and *Hyg* genes were used on the same PCR reaction. F1 and R2 generate an amplicon of approximately 600bp and F1 and R1 generate an amplicon of 400bp.

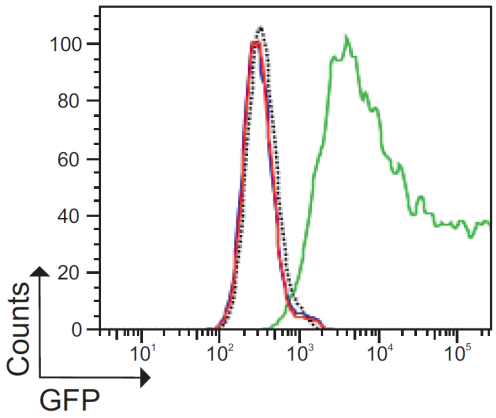


Fig. 36 – FACS analysis of 293T *PIGM*-promoter GFP recombinant cells. Cells were identified according to the physical properties SSC (side scatter) vs. FSC (forward scatter) and doublets were excluded by double gating on FSC-W (width) and FSC-A (area). Histograms shows GFP expression. Dotted line – unstained 293T cells; blue solid line – 910WT *PIGM*-GFP cells; red solid line – 910MUT *PIGM*-GFP cells; green solid line – 293T cells transiently transfected with the CMV-driven eGFP plasmid (positive control).

In order to verify whether GFP was in fact transcribed, qRT-PCR analysis was carried out. Under the control of the -270 WT *PIGM* promoter, cells expressed four times more GFP than cells harbouring the MUT promoter (Fig. 37A). This result was further confirmed at the protein level by Western blot analysis (Fig. 37B). Taken together these data suggested that GFP was expressed under the control of the *PIGM* promoter at low levels (as compared with the CMV driven plasmid).

Furthermore, the same result was obtained with another subset of WT and MUT cell clones (data not shown).

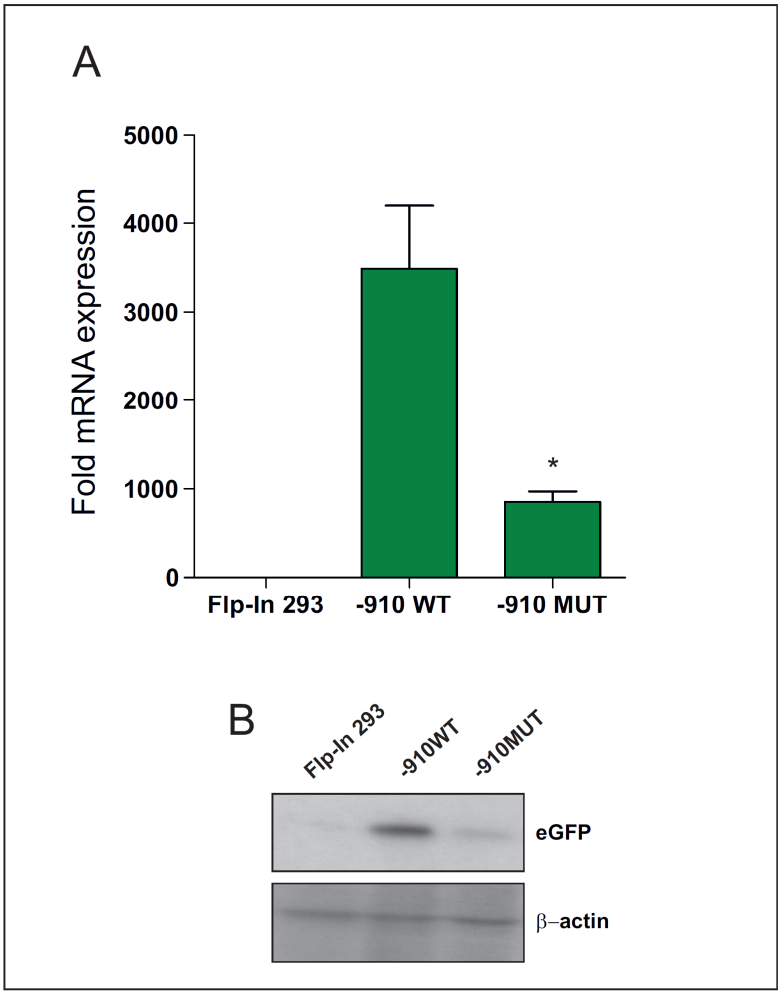


Fig. 37 – 293T *PIGM* promoter-GFP recombinant cells. A - *GFP* mRNA expression in parental Flp-In 293 cells, 910WT *PIGM*-GFP and 910MUT *PIGM*-GFP recombinant cells. *GFP* expression values were normalised to *GAPDH*. Results are presented as fold mRNA ± SEM of 3 independent extractions, ∗ $p<0.05$. **B –** GFP protein expression in the recombinant clones by Western-blot. β-actin was used as loading control.

Since *GFP* expression in 293 FRT recombinant cells recapitulated *PIGM* expression in WT and MUT B cell lines (see Results 3.1, Fig. 20A), overexpression of GATA-1 was finally performed in this cellular background. GATA-1 was previously sub-cloned in the MigR1-ratCD2 backbone vector and tested in HeLa cells by Western Blot (Appendix A, Fig. A8). Upon exogenous expression of GATA-1 in WT and MUT recombinant cells, GFP levels were assessed at 48h and 72 hours by qRT-PCR. Transfection efficiency was monitored by staining with rat-CD2 as exemplified in Fig. A9, Appendix A.

I found that upon GATA-1 transfection, *GFP* expression was comparable between WT and MUT cells, showing that the TF GATA-1 alone was not able to activate the *PIGM* promoter (Fig. 38). Whether in the presence of other erythroid-specific genes GATA-1 regulates the transcription of *PIGM* is not known.

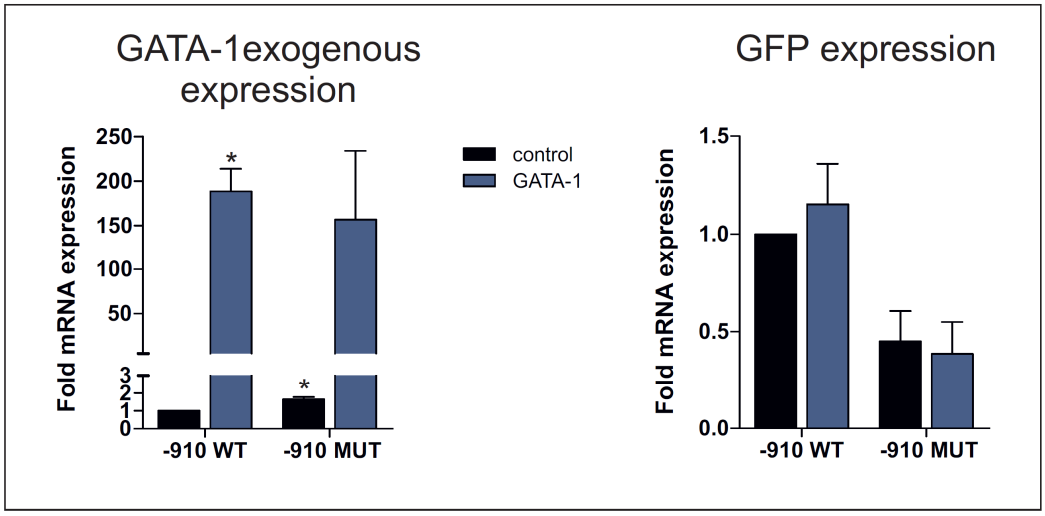


Fig. 38 – mRNA expression in – 293T *PIGM* promoter-GFP recombinant cells transiently transfected with MigR1-ratCD2 (control) or MigR1-hGATA-1 plasmids. 48 hours after transfection, cells were collected and assayed for mRNA expression by qRT-PCR. Left panel – *GATA-1* mRNA expression; right panel – *GFP* mRNA expression. Expression values were normalised to *GAPDH* and results are presented as fold mRNA expression (relative to control -910WT cells) ± SEM of 3 independent transfections, * $p<0.05$.

In summary, taken together the data suggested that the mega-erythroid lineage specific TF GATA-1 is not a transcriptional activator of *PIGM* and therefore could not account for active *PIGM* transcription in IGD erythroid cells.

KLF1 is a potential candidate in *PIGM* transcriptional activation

Given the crucial role of KLF1 as a transcriptional activator of erythroid genes, the next step was to investigate whether this TF contributes for the normal transcriptional regulation of *PIGM* in IGD erythroid cells. Thus, KLF1 binding profile at the *PIGM* promoter was analysed by ChIP-qPCR. Since KLF1 is not expressed in K562 (191), the analysis was carried out in primary erythroid precursor cells derived from PBMC (see Results 3.1.3, Fig. 14).

Whilst on Day 7 erythroblasts, KLF1 occupancy was comparable to background (2- fold enrichment), on Day 10, KLF1 showed a 4 and 5 fold enrichment at the -270 upstream region (amplicons 8 and 6) but not at the -270GC box (amplicon 4). Notably, a CACCC box, the binding motif of KLF1, lies in the region of amplicon 6, suggesting that KLF1 is a potential regulator of *PIGM* transcription in the erythroid cells.

Taken together, these results show that KLF1 is likely be an important transcriptional regulator of *PIGM*; however, its effect would be independent of the -270 motif.

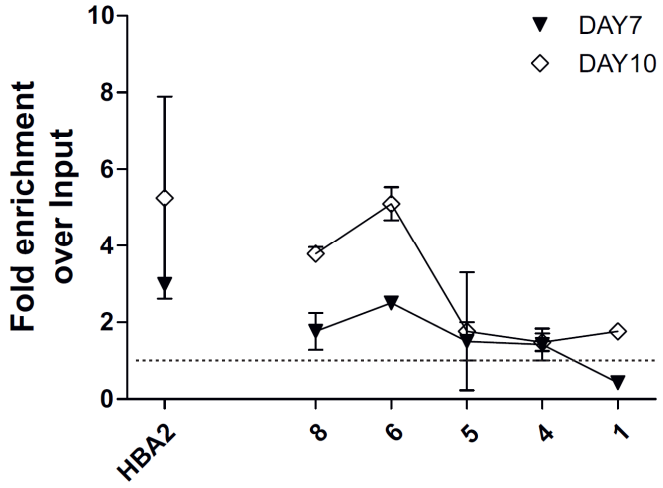


Fig. 39 – KLF1 occupancy at the length of the *PIGM* promoter in erythroid precursor cells differentiated *in vitro* from PBMC (see Results 3.1.3, Fig. 14). HBA2 locus was used as positive control. Dashed line – IgG control. IgG and KLF1 enrichment were expressed as % of Input. Results are presented as fold enrichment over input IgG ± SEM of 2 independent experiments.

4.1.4 Discussion

In *PIGM*-associated IGD, red blood cells have near normal expression of GPI despite the presence of the -270C>G mutation, indicating the existence of a differential mechanism of transcriptional regulation in relation to other blood cells such as B cells in which the majority are GPI-deficient (as showed in Results 3.1.3). To support *PIGM* expression in the presence of the -270C>G mutation, the erythroid master transcriptional regulators GATA-1 and KLF1 stood as two possible candidates.

Due to its short sequence, GATA-1 consensus motif occurs with high frequency in mammalian genomes (192, 193) and therefore, annotation of GATA-1 sites does not always predict GATA-1 binding *in vivo*. Indeed, various studies have shown that few WGATAR motifs at several loci are occupied *in vivo* by GATA-1 (194-197).

In order to define the role of GATA-1 as regulator of *PIGM* transcription, its binding profile was first assessed at the gene promoter. We showed that in the erythroid cell line K562, GATA-1 was modestly enriched at the predicted sites at the *PIGM* promoter (Fig. 31) and the overall enrichment in primary erythroid precursor cells at two different stages of differentiation was very low as compared to GATA-1 occupancy at the DNase Hypersensitive sites (DHSs) on its own locus (Fig. 32) and to IgG control. Despite the low GATA-1 enrichment at the predicted sites, it has to be considered that GATA-1 can occupy distal regulatory regions (198, 199) and therefore regulate transcription through long range interactions (200). Nevertheless, the data in primary erythroid cells needs further confirmation.

At the *PIGM* promoter two predicted GATA-1 binding sites (-712 and -675) lie within 50bp. Resolving TF occupancy at regions in such a close proximity constitutes one of the limitations of ChIP as the size of the immunoprecipitated fragments is usually within the range of 200bp. Nevertheless, new techniques derived from ChIP can bypass this limitation. For instance, ChIP-EXO is a technique based on the digestion of the sites flanking the TF-DNA complex with an exonuclease, followed by identification of the TFBS using deep-sequencing.

The advantages of this method are also related with the elimination of the uncrosslinked, unspecific DNA and with the achievement of resolution at single base level (201). Employment of this method or in alternative, performing *in vivo* mutagenesis of the nearby GATA-1 sites, could have provided better resolution in terms of GATA-1 occupancy at the promoter in K562 cells.

I could not investigate the GATA-1 binding profile in primary erythroid precursor cells derived from IGD patients due to the insufficient number of cells obtained after differentiation and the amount of chromatin required for this type of analysis (usually from at least $3x10^5$ cells).

In the erythroid cell line, GATA-1 occupancy is not responsible for the transcriptional activation of *PIGM*. Supporting this, a knockdown assay showed that GATA-1 strongly repressed *PIGM* as a modest reduction in 20% of the GATA-1 protein levels (Fig. 33C), increased *PIGM* expression by two fold (Fig. 34). In fact, it is known that GATA-1 not only participates in gene expression activation and consistent with its role as a repressor, it interacts with the NuRD complex, through FOG-1 (202, 203) and with the repressor complex LSD1-CoREST, via Gfi-1b (203, 204). Moreover, recent studies have suggested that in later stages of erythropoiesis, GATA-1 mediates transcription repression through PRC2 as GATA-1 repressed genes are also enriched for the repressive histone mark H3K27me3 (161, 199). Employment of an experiment using a second target shRNA and validation of the results in primary cells would help to better define the role of GATA-1 as a repressor of *PIGM* transcription.

As seen by others, in stable GATA-1 deficient cells, γ globin expression was reduced. In order to understand whether *PIGM* up-regulation results from a direct effect of GATA-1 knockdown, further studies are required. One possibility is the direct abrogation of the GATA-1 binding sites in the genome of the erythroid cell line K562 by using a genome editing technique such as the most recent system, CRISPR/CAS9 (205).

The finding that *PIGX* expression was also up-regulated, raised the possibility that GATA-1 mediated transcriptional repression extends to more genes of the GPI biosynthetic pathway. GATA-1 mediated transcriptional repression possibly requires the contribution of other repressive elements.

The use of transient reporter assays to study gene regulatory elements has fallen in disuse. Although very indicative of gene regulatory elements, these studies are often misleading due to the concentration effect of the several copies of plasmid introduced in the cell and the lack of the appropriate chromatin configuration. These intrinsic characteristics of the transient reporter assay frequently lead to aberrant activity. Generating a stable transfection assay in which the -270C>G mutation was site-specific integrated in the genome of the cells was definitely the most suitable approach.

In gain-of-function experiments in stable 293-GFP reporter cells, exogenous expression of GATA-1 was insufficient to activate the *PIGM* promoter and therefore to drive GFP expression as measured by qRT-PCR. *GFP* mRNA levels were comparable in WT and MUT cells, both highly expressing GATA-1. Whereas GATA-1 partners involved in gene repression are less clear, partners for gene activation have been well reported. GATA-1 mediated gene regulation often involves a combinatorial interaction with other transcription factors, such as KLF1 (206). Although GATA-1 can regulate gene expression independently of SCL/TAL-1 (161), gene expression mediated by the pentameric complex constituted by GATA-1/LMO2-LDB1-E2A/SCL-TAL-1 has been well studied (163, 164). More recently, other Krüppel-type zinc finger TF, ZBP-89, has been identified in TF regulatory complexes containing GATA-1 (207). All these factors are crucial either in early or late stages of normal erythropoiesis.

This being said, lack of activation of the *PIGM* promoter in stable 293-*PIGM* promoter cells may reflect the absence of the erythroid-specific, GATA-1 interacting factors in this cellular environment. Moreover, a controlled analysis of this system by testing the expression of GATA-1 target genes could not be assessed for the same reason.

Post-translational modifications constitute another important aspect of correct TF function. Accordingly, it has been seen that an acetylation defective GATA-1 mutant presents binding impairment *in vivo* (208) and recently, the importance of GATA-1 sumoylation has been demonstrated (209). Whether these modifications were correctly attainable in this system is not known.

Altogether these findings exclude GATA-1 as major activator of *PIGM* transcription in erythroid cells, raising the possibility that KLF1 has a key

regulatory role. Interestingly, evidence exists supporting that *PIG-Q*, an element of the enzymatic complex that transfers the N-acetylglucosamine to phosphatidylinositol (see Introduction, Fig. 4), is a KLF1 target gene (206). This reinforces the idea that master erythroid TFs regulate the expression of genes belonging to the GPI biosynthetic pathway thus ensuring the normal function of the erythroid cells.

Although the role of KLF1 could not be determined in detail, some evidence presented here suggest that it constitutes a potential activator of *PIGM* transcription, in a manner that is independent of the -270 motif and therefore could, at least in part, be responsible for intact *PIGM* transcription in IGD erythroid cells.

5.1 Exploring the role of the generic family of Sp transcription factors in *PIGM*-associated IGD

5.1.1 Introduction

In *PIGM*-associated IGD, the -270C>G mutation abrogates binding of the generic TF Sp1, causing selective histone hypoacetylation at the promoter of *PIGM*. Recently, we showed that in B cells, the pathogenic mutation that abrogates Sp1 binding to the -270GC-rich motif imposes nucleosomal compaction and mediates Polycomb-dependent repression of *PIGM* (38).

The generic TF Sp1, along with Sp2 and Sp3, belongs to the Specificity Protein/Kruppel-like Factor (Sp/KLF) family. Members of the Sp subgroup preferentially bind to DNA sequences highly enriched in GC-motifs. Although Sp1 was believed to regulate transcription of mainly housekeeping genes, studies showed that Sp1 can also regulate tissue-specific genes (165).
The function of Sp1 is modulated by its expression levels (210) , its protein-protein interactions and post-translational modifications (211), amongst others.

The lack of Sp1 binding, as a result of the -270C>G mutation, and its impact on the transcriptional regulation of *PIGM* has been clearly demonstrated in B cells, explaining their GPI deficient phenotype seen in IGD. By contrast, my studies in IGD erythroid cells suggest that the -270C>G mutation has no impact on *PIGM* mRNA expression and GPI surface levels; therefore, in erythroid cells, *PIGM* transcription is independent of the -270GC box. In addition, I ruled out transcriptional regulation by the erythroid lineage-specific TF GATA-1 as a mechanism that rescues *PIGM* transcription in IGD erythroid cells (see Results 4.1).

In this chapter I investigated **the role of Sp1 in *PIGM* transcription in normal and IGD erythroid cells**. For this purpose, I addressed the following: **i)** is *PIGM* transcription Sp1- dependent or –independent in erythroid cells; **ii)** what is the functional role of the several putative Sp1 binding sites (different to the -270GC-

rich box) at the *PIGM* promoter; **iii)** what is the role of the Sp-family members Sp2 and Sp3, which share the same DNA-binding domain with Sp1.

5.1.2 Experimental design

To dissect the role of the Sp family members in the transcriptional regulation of *PIGM*, I employed genetic (dominant negative Sp1-mutant), loss-of-function (shRNA) and pharmacological (MitA) assays. The binding profile of Sp1 and Sp2 was assessed by ChIP-qPCR analysis in cell lines and primary erythroid precursor cells.

5.1.3 Results

Sp1 regulates *PIGM* transcription in both erythroid and B cells

As employed in the knock-down (kd) assay described in Results 4.1, *Sp1* mRNA was specifically targeted by the use of shRNA lentiviruses in K562 cells and homozygous (N1) and heterozygous (2D) WT LBCLs. In this case, the Sp1 kd efficiency was first evaluated by qPCR analysis in transiently transfected 293T cells. shRNA Sp1 lentivirus was used in transduction experiments when at least a 50% reduction at the *Sp1* mRNA level was achieved in 293T cells (data not shown).

In the case of the WT homozygous LBCL, scramble- and Sp1 shRNA-transduced cells were sorted 96 hours post-transduction on the basis of GFP expression and were left to recover in the appropriate culture conditions.

I found that in Sp1 kd WT LBCL, *Sp1* levels were reduced to 35% of scramble control levels, commensurate with *PIGM* mRNA levels decreasing to 22% of control. Consistent with it being a Sp1-dependent gene (212, 213), dihydrofolate reductase (*DHFR*) mRNA levels were also drastically reduced (Fig. 40).

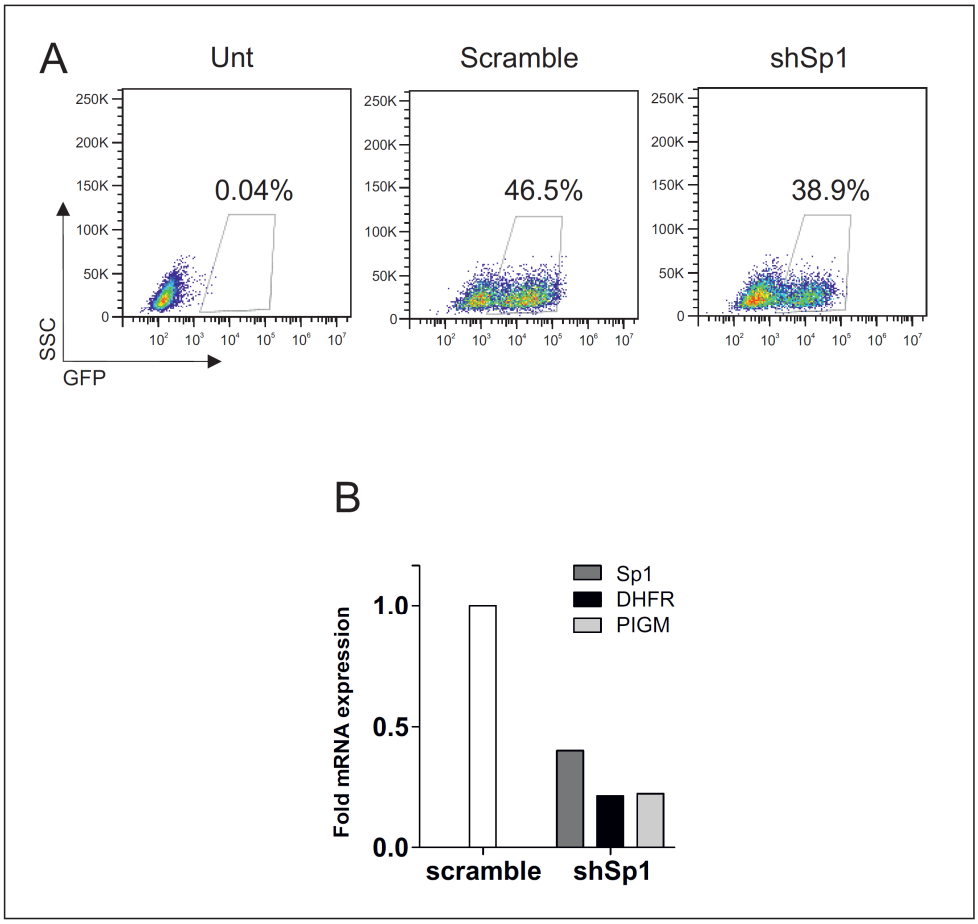


Fig. 40 – WT LBCL (N1) transduced with shRNA targeting Sp1 and control scramble. A – Cells were identified according to the physical properties SSC (side scatter) vs. FSC (forward scatter) of the cells and doublets were excluded by double gating on FSC-W (width) and FSC-A (area). FACS plots show GFP expression. Untransduced cells were used as unstained control sample. GFP+ cells were flow-sorted to high purity (>90%) **B** – mRNA expression in scramble and Sp1 kd WT LBCL. DHFR – positive control. Expression values were normalised to *GAPDH* and results are presented as fold mRNA relative to scramble cells, n=1.

In order to evaluate whether this effect was applicable to the heterozygous LBLCs, knockdown of Sp1 was also carried out in the -270C>G heterozygous cell line, 2D. After 96 hours post-transduction, *Sp1* mRNA levels were reduced to 13% of those in control scramble cells (Fig. 41). Consistently, *PIGM* mRNA levels and *DHFR* mRNA levels were approximately reduced by 60-70%.

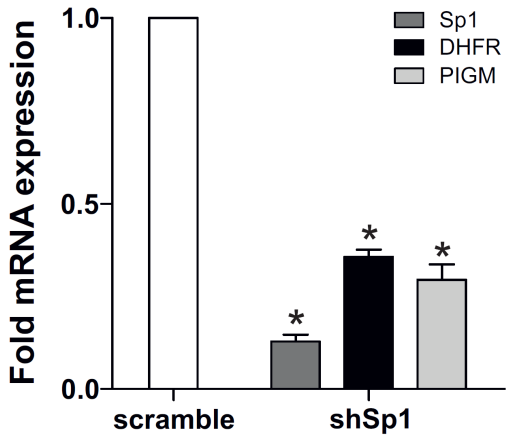


Fig. 41 – mRNA expression in the heterozygous LBCL scramble and LBCL Sp1 knock-down cells (LBCL-shSp1). DHFR – positive control. Expression values were normalised to *GAPDH* and results are presented as fold mRNA (relative to scramble cells) ± SEM of 3 independent transductions.

Taken together these results showed that in B cells *PIGM* transcription is activated by and depends upon Sp1.

Surprisingly, in the erythroid GFP-expressing K562 cells, knockdown of Sp1, performed in parallel with knockdown in B cells, was inefficient as seen by its mRNA levels following transduction with two independent target shRNAs (Fig. 42). Transduction efficiency was monitored by FACS analysis as described in Results 4.1, Fig. 33. The reason for this is unclear but it may be related, at least in part, to the 4-fold higher *Sp1* mRNA levels in K562 cells compared to B cells (Fig. 43).

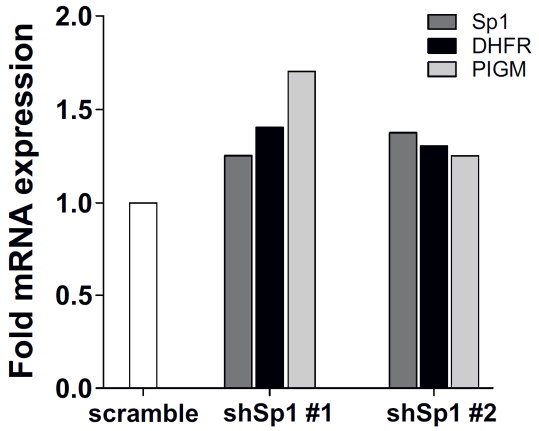


Fig. 42 – mRNA expression in K562 cells transduced independently with two shRNAs targeting Sp1. Cells were transduced with scramble, shSp1#1 and shSp1#2 lentiviruses and sorted 96 hours thereafter. Expression values were normalised to *GAPDH* and results are presented as fold mRNA relative to scramble cells, n=1.

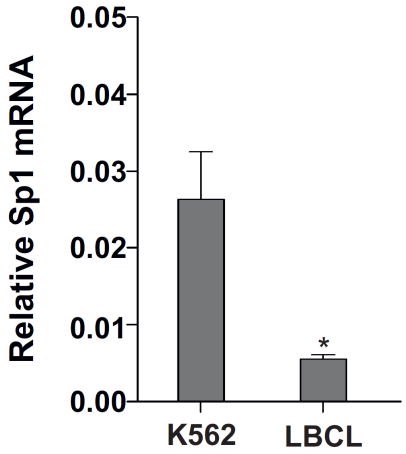


Fig. 43 – Sp1 mRNA expression in K562 and WT LBCL (N2) cells. Expression values were normalised to *GAPDH* and results are presented as mean ± SEM of 3 independent extractions, ∗ $p<0.05$.

In order to overcome this limitation, I employed an alternative method to interfere with the function of Sp1 and test its role in *PIGM* transcription in erythroid cells. Specifically, I used a dominant negative (DN) form of Sp1 which lacks the N-

terminus trans-activation domain but maintains its DNA-binding domain and therefore, once over-expressed in the cell, it competes with endogenous Sp1 for binding to its target genes (214). In addition, because the Sp1-mutant is not able to recruit the basal transcriptional machinery, transcription of the target genes decreases. After transferring this DN Sp1 cDNA from its original pEBGN-Sp1 backbone into the MigR-GFP vector (see Chapter 2, 2.12) I used it to transduce K562 cells.

After 48 hours post-transfection, GFP-expressing K562 cells were sorted and *PIGM* mRNA levels were analysed by qRT-PCR. By impairing endogenous, transcriptionally active Sp1 binding to its DNA cognate motifs, *PIGM* mRNA levels were reduced to 40% of its normal levels as was also the Sp1-dependent gene, *DHFR* (Fig. 44). These results show that just as in B cells, in erythroid K562 cells, Sp1 is also fundamental for *PIGM* transcription.

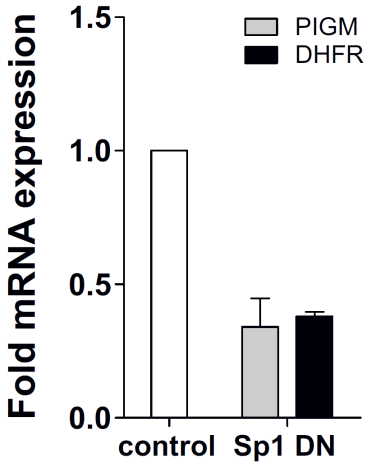


Fig. 44 – mRNA expression in K562 cells transduced with MigR1-GFP (control) or MigR1-Sp1DN-GFP retroviruses. Cells were sorted 48 hours after transduction. Expression values were normalised to *GAPDH* and results are presented as fold mRNA ± SEM of 2 independent transductions.

To further consolidate these findings, the erythroid and B cells were treated with Mithramycin A (MitA), an agent that inhibits TF binding to GC-rich motifs as are the TFs belonging to the Sp family (215-217).

After 24 hours of treatment at the indicated concentrations (0.5μM-1μM), cells were collected and assayed for *PIGM* mRNA and respective control genes.

By increasing MitA concentration in the -270C>G heterozygous LBCL 2D, *PIGM* mRNA levels decreased in a dose-dependent manner reaching one third of its normal levels at 0.4μM as compared to DMSO treated control cells. *DHFR* levels were reduced 5-fold of its normal levels and consistent with an auto-regulatory transcriptional circuit, *Sp1* expression was also drastically reduced (10-fold of normal mRNA levels) (Fig. 45A).

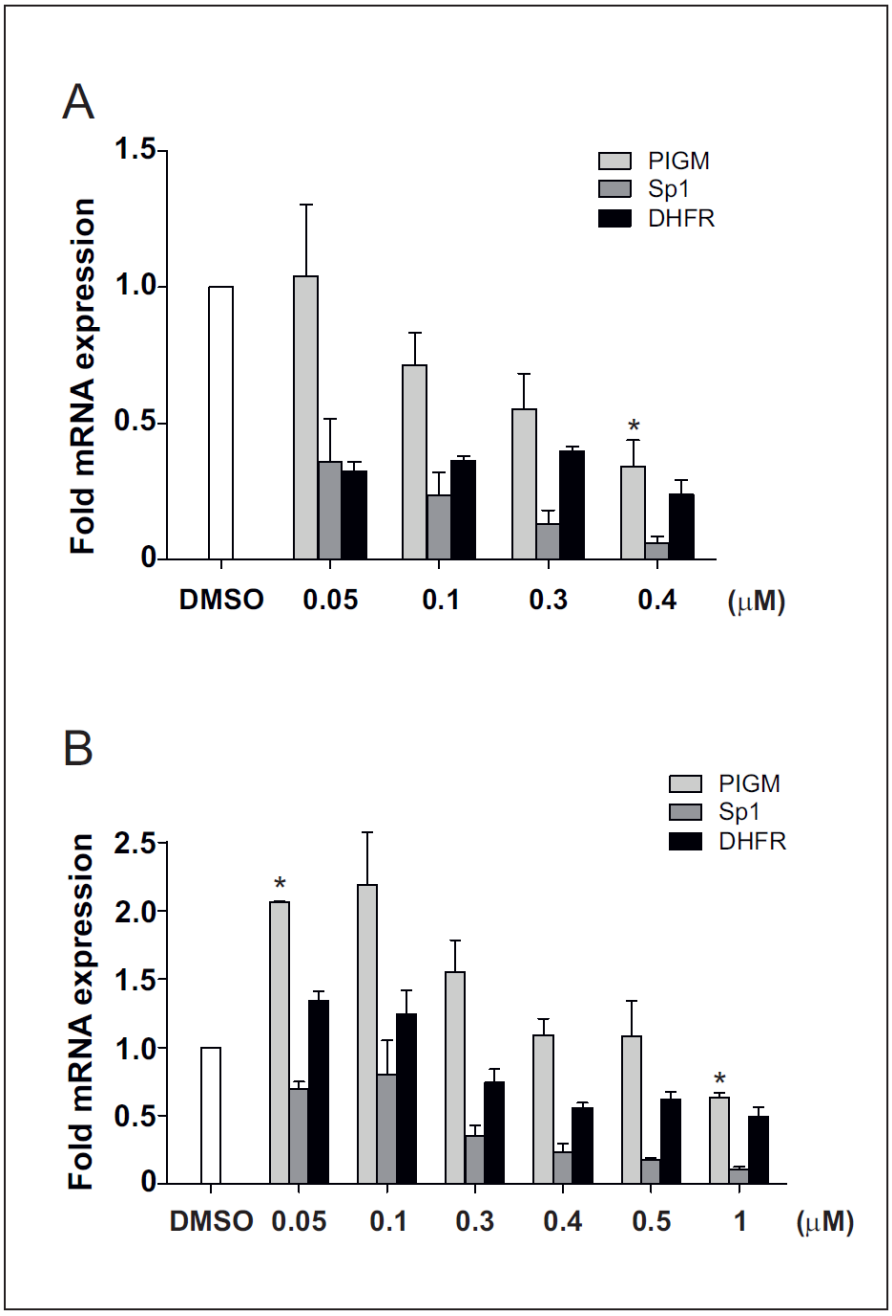


Fig. 45 – Effect of MitA on gene expression in the -270C>G heterozygous LBCL, 2D (A) and K562 cells (B) at the indicated concentrations, for 24hours. Expression values were normalised to *GAPDH*. Results are expressed as fold change mRNA (relative to DMSO-treated cells) ± SEM of 3 independent assays, ∗ $p<0.05$.

In the erythroid cell line K562, *PIGM* mRNA levels first increased at lower MitA (0.05 – 0.3μM) concentrations and then reduced to 40% at the highest concentration of 1μM. Regarding *Sp1* and *DHFR*, reduction of their expression

levels was achieved at the concentration of 0.3μM, to 35 and 74%, respectively (Fig. 44B).

Through the genetic (shRNA), loss-of-function (dominant negative Sp1-mutant) and pharmacological (MitA treatment) approaches the role of Sp1 in the transcriptional regulation of *PIGM* in the erythroid cells was better elucidated.
The investigations outlined here and in the previous section demonstrated that although Sp1 is required for *PIGM* transcription in erythroid cells, its effect is independent of the -270GC box.

As the *PIGM* promoter contains several GC-rich boxes, the next step consisted of evaluating the binding pattern of the Sp-family members by ChIP-qPCR.

Sp1 differentially occupies the GC-rich boxes at the *PIGM* promoter

Sp TF occupancy at the length of the *PIGM* promoter was determined by ChIP-qPCR using specific primers spanning the GC rich motifs.

In agreement with the results obtained in B cells, Sp1 occupancy at the *DHFR* promoter was used as a positive control whereas in the erythroid K562 cells, the *α–globin* locus was preferred (3). As reported in our previous work (38), Sp1 binds at the -270 GC-rich box in B cells (Fig. 46). A similar degree of Sp1 occupancy was seen at upstream promoter regions where at least one GC-rich box (-582; amplicon 6) was predicted. In the downstream -160 GC-rich motif (amplicon 1), Sp1 occupancy was comparable with background levels (fold enrichment of 2). Remarkably, in the erythroid cell line K562, Sp1 was 6-fold enriched at the corresponding upstream GC-rich box (amplicon 6) but not at the -270 GC-rich box (amplicons 4 and 3), suggesting differential, cell type-specific binding to this motif.
To further validate this finding, the same analysis was performed in primary $CD36^+$ erythroid precursor cells differentiated *in vitro* from $CD34^+$ cells isolated from cord blood. The advantage in using cord blood as source of $CD34^+$ cells is related with the high number of $CD36^+$ cells obtained after differentiation. As

seen by the expression of the surface markers CD71 and GlyA, $CD36^{+}$.cells were delayed in differentiation (Fig. 47A) when compared with cells derived from PBMC (Fig. 14). Nevertheless, ChIP analysis in the erythroid precursor cells also showed Sp1 binding selectivity, with Sp1 binding highly enriched at the upstream GC-rich box but not at the -270GC-box (Fig. 47B). The ChIP-seq data deposited at the ENCODE was not sufficient to distinguish with fine resolution Sp1 differential occupancy (see Appendix A, Fig. A10).

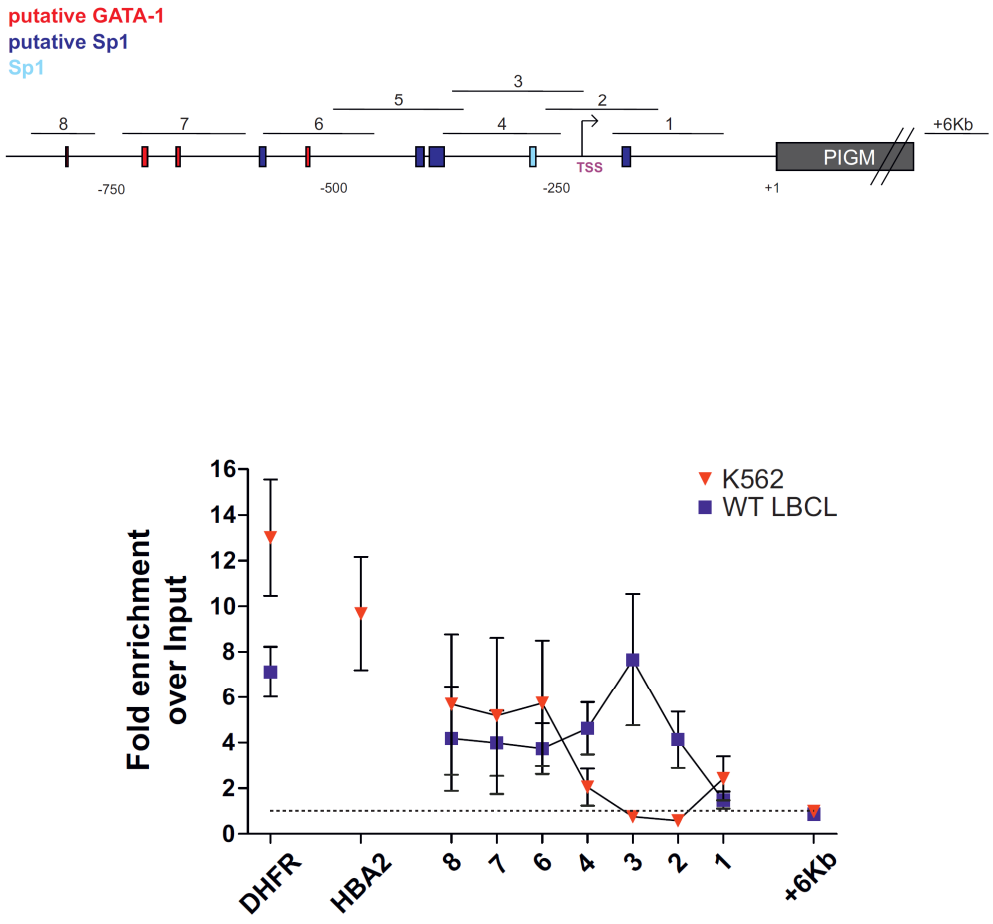


Fig. 46 – Sp1 occupancy at the length of the *PIGM* promoter in K562 and WT LBCL - N2. DHFR and HBA2 - positive control amplicons; +6Kb - negative control amplicon. Results are expressed as fold enrichment over Input ± SEM of at least 3 independent experiments.

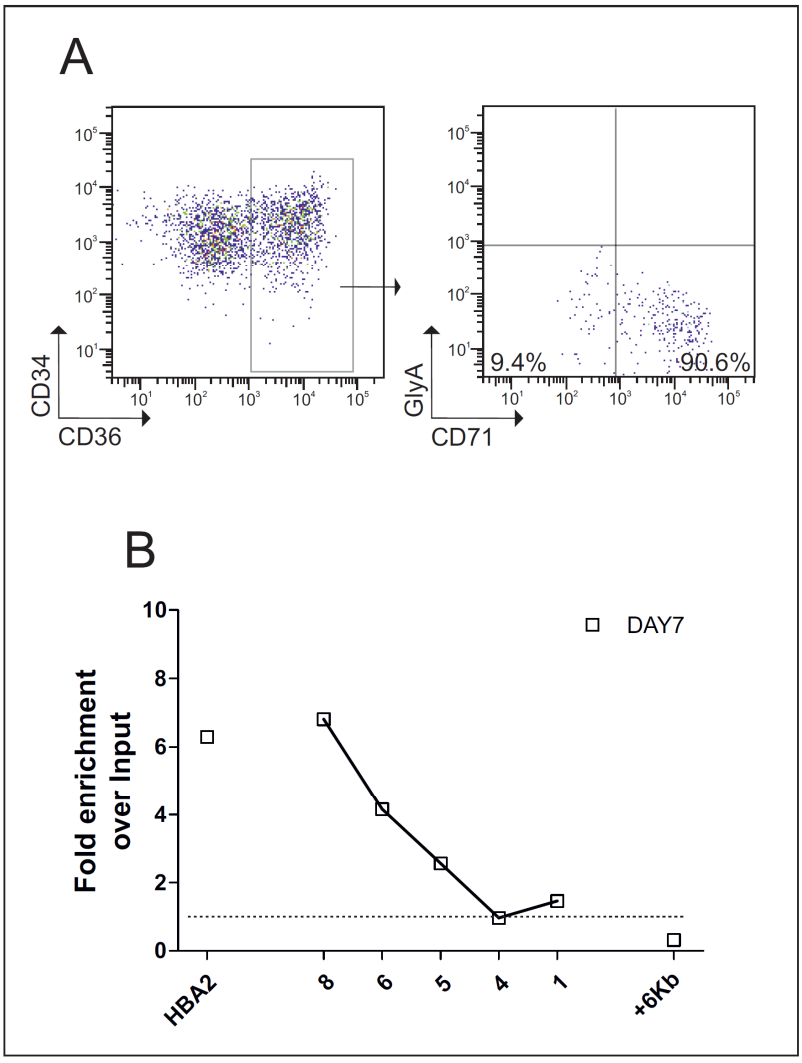


Fig. 47 – Sp1 occupancy at the length of the *PIGM* promoter in primary erythroid precursor cells. A – Cells were sorted according to the physical characteristics (FSC and SSC) and cell surface expression of CD34, CD36, CD71 and GlyA markers at day 7. **B –** Sp1 occupancy. HBA2 - positive control amplicon; +6Kb - negative control amplicon. Results are expressed as fold enrichment over Input, n=1.

Following these findings, the next step was to determine whether other members of the Sp family, namely, Sp2 and Sp3, could preferentially occupy the -270 GC-rich motif and therefore contribute for *PIGM* transcription and normal expression of GPI in patient erythroid cells. The results I obtained with the available anti-Sp3 antibody were inconsistent possibly due to it not being of sufficiently high quality for the purposes of ChIP assays. Nevertheless, the results I obtained in the ChIP of

Sp2 were very robust as seen by the high enrichment at the positive control region (Sp2 promoter), as described in (128). However, at the length of the *PIGM* promoter, including the -270 GC-rich box, Sp2 enrichment was uniform and slightly above the background (fold enrichment 2) in both WT LBCL and K562 (Fig. 48).

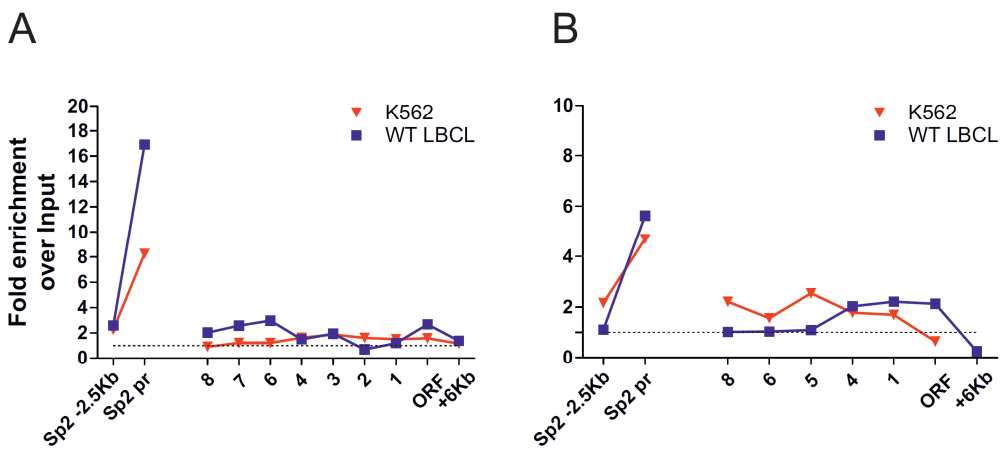


Fig. 48 – Sp2 occupancy at the Sp2 locus and at the length of the *PIGM* promoter by ChIP-qPCR in K562 (red) and WT LBCL - N2 (purple). Sp2 -2.5Kb and Sp2 pr (promoter) correspond to the amplicons used as negative and positive controls, respectively. Results from the two experiments (A and B) are expressed as fold enrichment over Input.

Altogether these results show that the general transcription factor Sp1 regulates *PIGM* transcription in erythroid and B cells. Whilst in B cells, *PIGM* transcription depends of Sp1 occupancy at the -270GC box, in the erythroid cells Sp1-mediated transcriptional regulation of *PIGM* is independent of the -270GC rich motif. Moreover, Sp2 does not directly contribute for the transcriptional regulation of *PIGM*.

Remarkably, these findings explain the normal GPI expression seen in patient erythroid cells, thus the lack of anaemia and intravascular haemolysis.

5.1.4 Discussion

A point mutation at the promoter of the mannosyltransferase-encoding gene *PIGM*, disrupts a GC-motif and abrogates binding of the TF Sp1.

The role of Sp1 in the transcriptional regulation of *PIGM* was first dissected by employing a knockdown assay. While in the B cells, the *Sp1* mRNA levels were significantly reduced (Fig. 41), in the erythroid cells, Sp1 levels were comparable to those in control cells transduced with a scramble lentivirus (Fig. 42). The impossibility in generating Sp1—deficient K562 cells by shRNA might be the result, at least in part, of the higher levels of mRNA in these cells (Fig. 43). Failure to target by the two shRNA an Sp1 isoform that might be abundant in K562 cells might be another possible explanation for the inefficient knockdown of Sp1 in these cells.

In contrast to K562 cells, in WT and heterozygous -270C>G B cells, efficient reduction of Sp1 drastically decreased *PIGM* expression.

Regardless the inefficient Sp1 kd in the erythroid cells, the role of Sp1 in the transcription of *PIGM* was efficiently demonstrated by overexpressing a dominant negative Sp1-mutant. Accordingly, in the presence of Sp1 lacking of transactivation potential and failing in the recruitment of the basal transcriptional machinery, *PIGM* expression was reduced to levels lower than 50% (Fig. 44).

Another manner to test Sp1dependency is by employing a pharmacological approach using Mithramycin A (MitA). MitA has been shown to bind to GC modules and thus inhibiting Sp binding, including Sp1, to the DNA and repressing gene expression (216). Diminished Sp1 DNA binding to GC-boxes by using MitA caused a reduction in the *PIGM* levels (Fig. 49). However, a higher concentration was required in erythroid than in B cells (1μM vs 0.4μM, respectively). Again, this might be related with the high levels of Sp proteins in the erythroid cells and therefore the requirement of more MitA for binding competition.

Surprisingly, at lower doses of MitA in the erythroid cells, *PIGM* and *DHFR* expression increased whereas *Sp1* expression decreased, suggesting the involvement of other TFs in the erythroid cells in counteracting the role of the Sp TFs or alternatively, suggesting that members of the Sp family repress *PIGM* (and *DHFR*) transcription.

Altogether these data showed direct involvement of Sp1 in mediating *PIGM*-transcriptional regulation in both erythroid and B cells.

Unexpectedly, the data showed that in erythroid cells, Sp1 was preferentially recruited at upstream GC boxes rather than at the -270GC box (Fig. 46). In B cells, Sp1 occupancy was high and extended at the length of the promoter with highest binding levels at the -270GC rich box, highlighting the importance of this motif in *PIGM* transcription in B cells. Recent *in vitro* mutagenesis studies have suggested that, in the context of the multi-functional *TNIP1* gene, when multiple, nearby Sp1 sites are available for binding, the gene is more prone to repression as a result of repressive TFs forming a complex with Sp1 (218). In the context of *PIGM*, Sp1 binding at the length of the promoter in B cells and relative to other cell types, is indeed associated with a relative higher level of nucleosomal compaction configuration that leads to reduced transcription, as seen in Results, 3.1.3. By contrast, in the erythroid cells, exclusive Sp1 binding at the upstream GC box results in a more open, nucleosome-free promoter configuration that ensures higher levels of *PIGM* transcription.
ChIP-seq data deposited in the ENCODE database showed low resolution in terms of Sp1 binding at such nearby sites and therefore could not further support this finding. As mentioned before (see Results 4.1.3), employment of ChIP-EXO could have bypassed this limitation.

Studies showing cell-type specific preferences for a given TF at the length of the same gene promoter have not been reported. The closest investigations come from a study in which a model to predict differential TF occupancy within a single-cell type was generated. Chromatin accessibility was not predictive of TF binding, as in regions where TFBS were accessible in both erythroid and B cell lines, TFs presented cell-type preferential binding. The study showed that TF binding has to be considered in the context of specific nucleotide sequence, heteromeric complex formation, post-translational modifications, and other allosteric alterations (39).

In vitro studies have suggested that Sp1 competes with Sp3 for the same binding site (219, 220) and other studies have demonstrated that Sp1 and Sp3 selectively

bind to different Sp1/Sp3 site(s) on the same promoter, depending on the spatial availability (221). By recruiting different complexes, Sp1 and Sp3 can exert an active or repressive effect in transcription (222). If there is Sp1/Sp3 selectivity at the *PIGM* promoter could not be determined. Although conclusions can not be drawn, the possibility that Sp3 binds to the -270GC-rich motif, thus contributing to the normal expression of *PIGM* in the IGD erythroid cells, can not be excluded. However, my studies clearly show that Sp2 is not a direct regulator of *PIGM* transcription.

Speculations apart, depicting unknown TFs bound at the gene regulatory regions can be achieved by reverse-ChIP techniques such as Proteomics of Isolated Chromatin segments (PICh) (223) or similar, like Chromatin Affinity Purification with Mass Spectrometry (ChAP-MS) (224). The former involves isolation of the protein –DNA complex by using a nucleic acid hybridisation method followed by mass-spectrometry. The nuclei acid probe targets the genomic locus of interest and subsequently all the proteins bound to that locus are identified according to their mass to charge ratio. Employment of this method would have allowed the identification of the TF(s) bound at the -270GC motif in the erythroid cells and therefore the study of their role on *PIGM* transcription.

At the most proximal -160GC-box I found that Sp1 was not enriched in both erythroid and B cells (Fig. 46). By prediction, mutations in this putative Sp1 binding site have no effect in *PIGM* transcription and therefore in GPI expression in these cells.

In this chapter, I showed that Sp1 is required for *PIGM* transcription in erythroid and B cells. Very importantly, I demonstrated that Sp1 discriminates between GC-boxes at the *PIGM* promoter in a cell-environment dependent manner. These finding explains the normal *PIGM* transcription and GPI expression in patient's erythroid cells and consequently the lack of anaemia and intravascular haemolysis.

6.1 Investigating the genomic interactions of *PIGM*

6.1.1 Introduction

In the nuclear space, chromatin is not randomly distributed but presents instead a level of organization. Chromosomal territories (CT) represent the topological space occupied by chromosomes, are the highest-order of chromatin organization and can be cell-type specific (225).

Preferential organization of the genes in the nuclear space has been proposed to associate with their transcriptional status. Accordingly, there are regions in the nuclear space where transcription is usually active (euchromatin regions like in the nucleolus) and others where it is not (heterochromatin regions). Transcriptionally active genes may loop out of their CT to create "hubs" of transcription or transcriptional factories, and come in close proximity of other in *cis* or in *trans* genes with which they interact. In transcriptional factories, are transcriptional units in which participating genes share of the use of TFs and active RNA PolII, that are essential for active gene transcription (226). Further studies in erythroid cells showed that *in cis* active genes co-localise in foci of active transcription, and once moved out transcription is abolished. Recruitment into the factories is most likely influenced by effectors of transcriptional regulation such as TFs, regulatory elements (enhancers) and epigenetic marks (227).

This fascinating structural and functional organisation is supported by the findings that transcriptional regulation does not occur in a linear fashion and the regulatory elements do not necessarily regulate transcription of the nearby genes. Indeed, regulatory elements like enhancers contact the gene proximal region of the target gene by TF-DNA-mediated looping formation.

DNA looping formation and the three-dimensional organisation of the genes in the nucleus is therefore relevant for transcriptional regulation and constitutes a layer of regulation that should be taken into account.

Long-range interactions can be detected and quantified by using the chromosome conformational capture (3C) methodology and its derivatives. The technique is based on the detection of intramolecular ligations following crosslink of the chromatin. The abundance of a given intramolecular ligation reflects the frequency of an interaction and therefore, the vicinity between two genomic loci (4). Circular chromosome conformational capture (4C) is a derivative of 3C. Whereas in the 3C both bait and interactor are known, in the 4C the approach is unbiased. 4C is based on the proximity ligation principle, in which interacting protein-DNA complexes are favoured and circularize under diluted conditions; it involves two steps of restriction digestion with appropriate enzymes and identification of the interacting genes by genomic screening (microarray analysis or high throughput-sequencing). Regarding the choice the restriction enzyme, some aspects have to be taken into account. The choice of a 4 or 6bp cutter depends of the size of the loci under analysis; the enzymes have to efficiently digest the chromatin at 37°C as high temperatures uncross-link the protein-DNA complexes; the enzymes that are sensitive to DNA methylation should be avoided as they may introduce a bias into the assay due to different DNA digestion efficiency. The 4C assay presents higher resolution than 3C and is highly accurate in the quantification of DNA interaction frequencies of a specific genomic area (228).

In the previous chapters I elucidated the role of a *cis*-DNA element, in this case of the -270GC-rich box, in differential, tissue-specific transcriptional regulation of *PIGM*. In an attempt to better understand **the complex regulatory network of *PIGM*, here I propose to investigate whether *PIGM* established in *cis* or in *trans* long range interactions with other genes as part of the same transcriptional unit. Furthermore, I intend to investigate the impact of lack Sp1-binding at the *PIGM*-promoter in IGD cells in this process.**

6.1.2 Experimental design

In order to identify *PIGM*-interacting genes, 4C assay was employed in uncross-linked (controls) and cross-linked WT- and MUT (IGD)- LBCLs followed by

high-throughput sequencing. Bioinformatic analysis was performed by the Computational Genomics Analysis and Training (MRC Functional Genomics Unit, Oxford).

The results presented below are still preliminary.

6.1.3 Results

PIGM and neighbour genes are part of a transcriptional unit

Following previously optimization, ApoI (6-bp cutter, R^AATTY) and NlaIII (4-bp cutter, CATG^) were the restriction enzymes of choice. Following sequential digestion intercalating with two steps of intramolecular ligation, the libraries for the WT and MUT cells were prepared (see Chapter 2, 2.6 and schematic representation below). Controls included digestion efficiency by qRT-PCR and specificity of the PCR amplification by DNA sequencing.

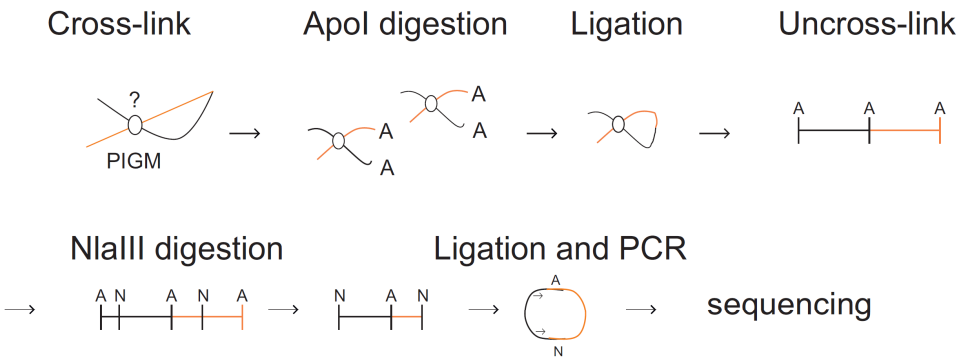


Fig. 49 – Schematic representation of the 4C assay. Following digestion and intramolecular ligation, PCR amplification of the fusion sequences is performed by using tail to tail *PIGM* specific primers. Adapted from (228).

Ligation junctions were captured by high-throughput sequencing and data was aligned to the reference genome and then compiled as heat-maps. The specificity

of a given interaction is supported by the identification at the fusion of the partner genes of the ApoI/NlaIII sites. Therefore, interactions are revealed as fusion sequences between the bait, i.e., *PIGM* and the interacting partner.

In the WT uncross-linked cells (Fig. 51B), the majority of the *PIGM* interactions were localised at the *PIGM* locus, most likely as a result of re-ligation. In the cross-linked samples (Fig. 51A), high intensity contacts (seen as peaks) were established with the following neighbour genes: *IGSF9*, *SLAMF9*, *KCNJ10, KCNJ9*, *IGSF8* and *ATP1A2*. Less intense contacts were established with *CASQ1*, *DCAF8* and *NCSTN* genes.

Regarding the MUT cells, in the uncross-linked state (Fig. 51D) the number of captures was less than in the WT uncross-linked cells but still at the *PIGM* locus (overall MUT cells were somehow more resistant to digestion). In the MUT cross-linked sample (Fig. 51C), the 4C profile was similar to the MUT UNX or WT UNX, suggesting inefficient cross-link. However, as uncross-linked samples were less resistant to digestion than cross-linked samples, this indicated that the MUT cross-linked profile represented a real loss of interactions and it was not the result of inefficient cross-link.

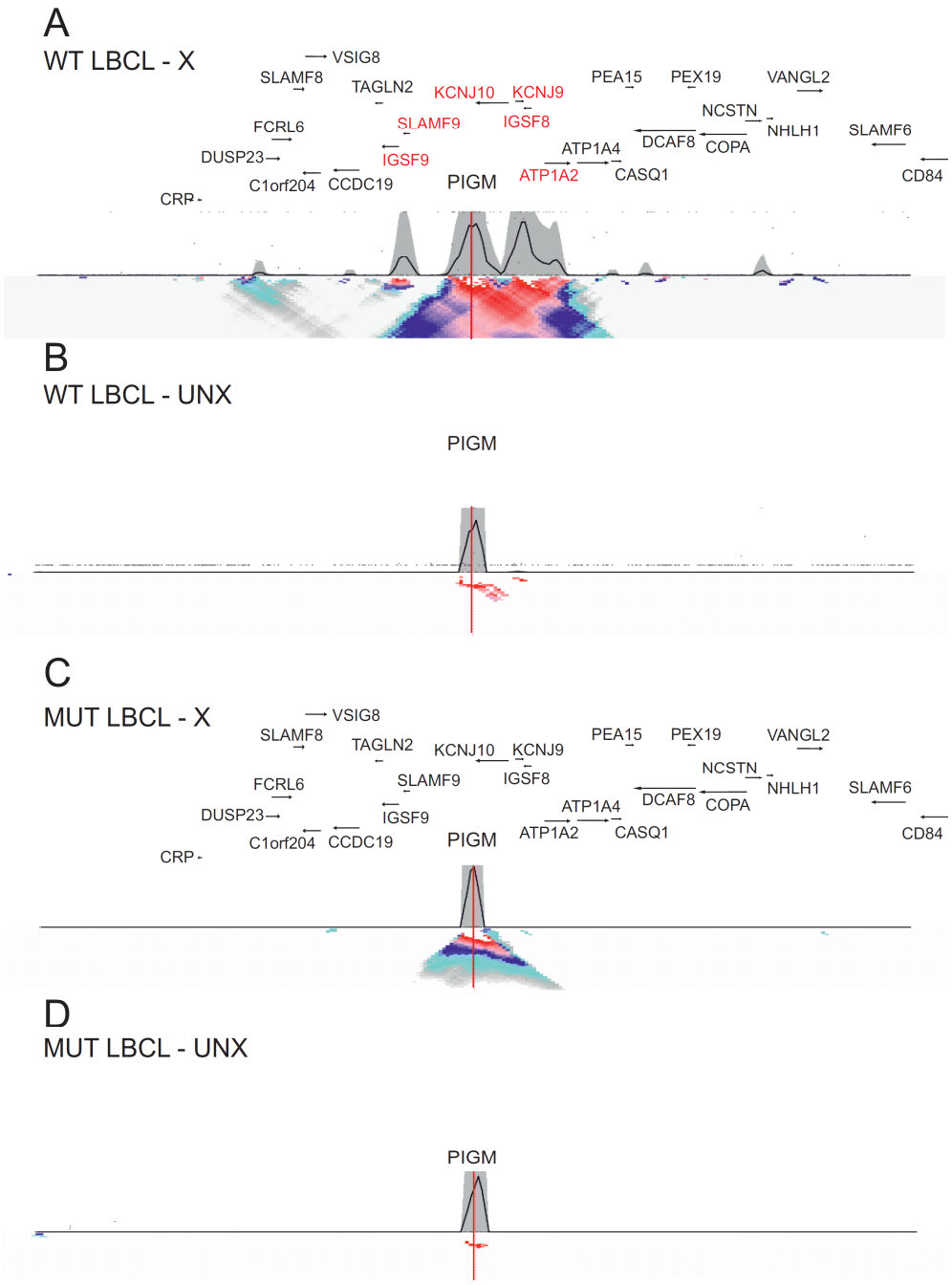


Fig. 50 – Heat map showing the average contact intensities of *PIGM*. A – cross-linked and **B –** uncross-linked WT LBCL (N2); **C –** cross-linked and **D –** uncross-linked MUT LBCL (2C). Window size varies from 4Kb to 200Kb.The colour scheme (grey to red) represents the frequency of the contact intensities.

The results in B cells suggest a common regulatory network shared between *PIGM* and the interacting partners, as part of a transcriptional unit. Given that in

the MUT cells the contacts are lost and preliminary data showed alignment of the interactions with known Sp1-binding sites, active gene expression, promoter state and open chromatin (ENCODE), the results further suggest that Sp1 could be involved in mediating the *cis*-interactions by DNA-loop formation.

6.1.4 Discussion

PIGM-associated IGD is a disorder characterised by abdominal vein thrombosis and epilepsy. The complexity of the disease translates the diverse GPI phenotype seen in patient cells, as a result of differential transcriptional regulation of *PIGM*. In B cells, *PIGM* transcription is dependent of Sp1 binding to the -270GC box. Using as a model LBCLs derived from a patient and a normal donor B cells, I tested whether *PIGM* is part of a transcriptional unit, sharing with genes in the close proximity similar mechanisms of regulation and whether in the absence of Sp1 binding to the *PIGM* promoter and lack of active transcription, the regulatory network where these genes are embedded, is disrupted.

Preliminary data resulting from 4C analysis showed *cis*-interactions established between *PIGM* and the following genes: *IGSF9*, *SLAMF9*, *KCNJ10, KCNJ9*, *IGSF8* and *ATP1A2*. Those interactions were captured with high frequency, suggesting that in the nuclear space these genes are indeed in close proximity with *PIGM*.

KCNJ9 and *KCNJ10*, both part of a subfamily of inward rectifier potassium channels constitute two of the high-frequent *PIGM*-interacting genes. Interestingly, SNPs in the region encompassing the *KCNJ9*/*KCNJ10* genes have been associated with temporal lobe epilepsy (229), suggesting their role in the maintenance of the brain function. It is very tempting to speculate that lack of appropriate *PIGM* transcription can influence their expression, constituting an explanation for some of the IGD characteristics (such as epilepsy). Furthermore, *IGSF8* and *ATP1A2*, two other genes depicted from the analysis, have been suggested to participate in the maintenance of the neural network in the adult brain. *ATP1A2* is associated with a rare neurological disorder that is Alternating

Hemiplegia of Childhood (AHC), characterised by recurrent attacks of hemiplegia (230, 231). *SLAMF9* is involved in the signalling lymphocytic activation (232).
In view of the functional aspects of the *PIGM*-interacting genes the results point towards a regulatory link between these genes and *PIGM* that once disrupted possibly contribute for the phenotype of the disease. Further validation of the ligation junctions would require 3C analysis by using primers targeting both *PIGM* and the interacting gene as well as DNA-FISH, using specific DNA-probes to detect co-localization/proximity. Determining whether the captures are functional by quantifying gene expression using RNA-seq is also essential. Lastly, in order to consolidate these findings occupancy of Sp1 at H3K4me marked promoter regions of the interacting genes needs to be carried-out by ChIP. Those results will reinforce the role of Sp1 as an "anchor" in mediating *PIGM*-loop interactions.
Furthermore, I speculate that NaBu treatment of the MUT cells restores the interactions seen in the WT cells.

Although preliminary, these findings suggest that *PIGM* is part of a transcriptional unit in which genes important for the brain function are involved, possibly sharing a common mechanism of Sp1-mediated transcriptional regulation. Sp1-mediated chromatin looping has been reported in the enhancer-promoter *HO-1* (heme oxygenase 1) gene in renal cells (129) .

Altogether these results add another level of complexity at the *PIGM* locus and contribute for the knowledge of such a particular disease. Further, they reveal the importance in identifying genes contained in transcriptional units as a unique contribution for the better understanding of disease.

Chapter 4 – Discussion and concluding remarks

Appropriate regulation of gene expression is essential for the maintenance of the cellular function. In recent years, genome-wide studies have delivered an enormous amount of data in relation to the structure and function of regulatory elements and mechanisms of gene transcription. This data mostly results from high-throughput studies including RNA-seq, ChIP-seq, DNase-seq, Hi-seq and others. From the integration of gene expression data, TF occupancy, chromatin accessibility and epigenetic marks, novel regulatory elements have been identified (29) suggesting the presence of even more complex regulatory interactions within the mechanisms of gene regulation.

However, these tools become less useful when studying regulation of gene expression at the level of a single gene and particularly in a cell-type specific context.

PIGM-associated IGD constitutes a very interesting model to study the impact of a cis-acting mutation in gene expression (de)regulation at the cell- and tissue-specific level. Cis-regulatory mutations are usually screened after coding mutations have been excluded. In striking contrast to the other genes required for the GPI biosynthesis, *PIGM* is so far the only gene of the pathway in which a functional mutation in a regulatory region has been reported. This may be related with the fact that mutations are not always compatible with life and therefore create a "gap" in the number of studies available. Interestingly, inherited mutations in other genes of the GPI biosynthetic pathway are coding and are associated with a more severe phenotype that in *PIGM*-associated IGD. This may be related to the fact that regulatory mutations are often hypomorphic, i.e., do not totally abrogate expression and even if low, the amount of protein/enzyme that is produced is functional and can be sufficient in a certain cellular environment to reach the required threshold. On the other hand, a coding mutation, even if it is missense, may produce an aberrant and/or completely dysfunctional protein.

The -270C>G pathogenic mutation disrupts binding for the generic TF Sp1 in the core promoter of *PIGM*. As we recently published, impaired TF binding is associated with selective histone hypoacetylation, nucleosomal compaction, Polycomb recruitment and imposition of bivalent histone methylation marks. These events ultimately lead to a down-regulation of *PIGM* mRNA and deficient GPI expression at the cell surface in IGD B cells. Contrasting with PNH, another well-characterised disorder of the GPI pathway, GPI expression at the surface of the erythroid cells is near normal thus the absence of anaemia and intravascular haemolysis in patients. Similarly, in the group of Inherited GPI disorders, patient erythroid cells possibly express normal levels of GPI as deficiency has not been reported. As within this group most disorders result from coding mutations I hypothesize that in those where the catalytic domain of the enzyme is affected, the deficient enzyme is somehow re-activated in the erythroid environment or its activity is sufficient for minimal expression. The molecular events driving normal GPI expression in the erythroid cells of patients with *PIGM*-associated IGD were not known and constituted the overall aim of this thesis.

Sp1 may discriminate between GC-boxes along the promoter of *housekeeping* genes

In *PIGM*-associated IGD, GPI deficiency results from transcriptional deregulation of *PIGM*. For instance, in the B lymphocyte lineage GPI- cells express less *PIGM* mRNA than GPI+ cells (Fig.19).

In WT B cells the *PIGM* promoter architecture is more compacted than in the erythroid cells, and therefore in the former the gene is more prone to transcriptional repression (Fig. 23A and B). Associated with this, Sp1 binds along the GC-boxes at the promoter, most highly at the -270GC box in B cells (Fig. 46).

By contrast, in erythroid cells, the pathogenic mutation had no impact at the transcriptional level, as mutant primary erythroid precursor cells expressed *PIGM* mRNA (Fig. 19) and the GPI-linked antigen CD59 (Fig. 18B) as compared to their heterozygous counterparts. Moreover, in the erythroid cell line the TSS is positioned downstream at the -270GC motif (Fig. 25), excluding the possibility of

transcription initiation upstream of the pathogenic mutation and differential core promoter *usage*.

The most striking finding of my investigations was the absence of Sp1 binding at the -270GC-box in the erythroid cell line K562 (Fig. 46) further confirmed in primary erythroid precursor cells (Fig. 47). This result revealed that Sp1-mediated transcriptional regulation of *PIGM* in erythroid cells is independent of the -270GC box, thus explaining the normal GPI+ phenotype seen in patient mature red cells.

Although numerous studies have focused on cell type-specific gene regulation, there is a lack of investigations aiming to dissect cell type-specific TF binding preferences at gene regulatory regions. The studies reporting differential, cell type-specific TF preferences are usually focused in motif-comparisons between genes and not on motifs on the same gene (34, 39, 40). Based on the results of my studies in normal and IGD erythroid cells I propose that discriminative binding to GC boxes is the result of an erythroid cell-specific configuration of the *PIGM* promoter (see proposed model, Fig. 51). Accordingly, protein-protein interactions established upstream at the promoter between erythroid specific TFs (possibly KLF1) may limit the spatial availability and by nucleating the transcriptional machinery potentiate *PIGM* transcription. Whereas KLF1 is a possible direct *PIGM* transcriptional activator, GATA-1 is most likely involved in *PIGM* repression.

Another possible scenario might involve reduced Sp1 binding affinity to the -270GC box possibly due to those interactions established upstream at the promoter, so altering Sp1 specificity or sequence affinity. Additionally, investigating how the levels of Sp1 and its post-translational modifications affect Sp1 availability and binding at the promoter in the erythroid cells could have been important to better dissect the regulatory mechanism driving *PIGM* transcription.

In addition, *PIGM* mRNA levels in the erythroid cells suggested a low basal level in normal conditions as compared to the B lymphocytes (Fig. 15). Nevertheless, the amount of protein that is produced is sufficient for normal enzymatic function and therefore for the normal cellular processes. Comparing PIGM activity in the eryhroid cells with that in the B cells would contribute for the better understanding of the differential threshold and as a consequence, of the disease.

Thus, I speculate that the role of the -270GC box is to potentiate *PIGM* transcription in cells where higher levels are needed, like in B cells.

In patients with *PIGM*-associated IGD, cells belonging to the monocyte lineage as well as T cells expressed normal GPI levels whereas the B lymphocytes were mostly deficient (Fig. 17). Deciphering the regulatory mechanism of *PIGM* in these cell lines is beyond the scope of this thesis. Whether in the monocytes and T cells, *PIGM* expression is independent of the -270GC box, is not known. Nevertheless, I speculate that in the monocytes, in which the *PIGM* basal levels are also low, Sp1 is not recruited at the -270GC box.

An interesting observation is that the erythroid precursor cells contrast with the red blood cells (RBC) isolated from peripheral blood of patients in respect to CD59 expression. While the *in vitro* differentiated erythroid precursor cells express normal levels of CD59 (Fig. 18B), the red blood cells isolated from peripheral blood present some degree of heterogeneity, with some cells expressing less CD59 (Fig. 13). This might be related with gene expression shut-down in RBC combined with the fact that in the erythroid culture system used in here cells were collected at the proerythroblast stage and the *PIGM* mRNA levels were still elevated.

By disrupting erythroid-specific transcriptional networks, Sp1-binding mutations have been shown to affect gene expression in erythroid cells (93, 139). Here we show that the Sp1 binding mutation is not disadvantageous to the erythroid cells as a result of lack of Sp1 binding in the normal conditions. Additionally, the complexes between generic and erythroid-specifc TFs recruited upstream at the promoter are most likely fundamental for normal expression.

In view of the importance of generic TF binding in *housekeeping* gene regulation, understanding whether Sp1 discriminates between GC-boxes at the regulatory regions of other *housekeeping* genes, usually enriched in GC-motifs, would doubtless provide invaluable information. As regulatory mutations not always underlie an evident phenotype and might have an impact at the cell-type specific level, this link may be difficult to study in a genome-wide type of approach.

I speculate that this finding is applicable to other *housekeeping* genes. Discrimination between GC-boxes might constitute a mechanism of regulating different threshold levels in different cell types, thus always ensuing sufficient mRNA levels.

The -270GC box may be bound by TFs other than Sp1

Having seen that KLF1 (erythroid TF that shares the DNA binding domain with Sp1 with similar affinities) was not recruited at the -270GC box, but it was instead recruited at the upstream promoter region of *PIGM* (Fig. 39), I hypothesise that in the erythroid cells, *PIGM* transcription is most likely driven by a combination of generic and cell-type specific TFs, like Sp1 and KLF1, respectively, anchored at the proximal promoter upstream at the -270GC box. Evidence exists supporting that Sp1 and KLF1 synergistically interact to regulate gene expression of enzymatic proteins (233). Moreover, a gene of the biosynthetic GPI pathway (*PIGQ*) is a KLF1 target in the eryhroid cells (206), suggesting that erythroid specific TFs regulate expression of *housekeeping* genes belonging to the GPI pathway.
Whether transcriptional regulation of *PIGM* involves recruitment of the upstream regulators nearby the TSS by DNA-looping formation is not known. Studies supporting the existence of short-range interactions in mediating gene regulation have been reported (234).

The role of Sp3 in competing with Sp1 for the DNA-binding domain can not be excluded. Also Sp4, mainly expressed in neuronal cells, may contribute for the regulatory mechanism driving *PIGM* expression in the neuronal tissue of the affected children. The importance of Sp proteins in the regulation of neuronal genes has been reported (101, 235). Specifically, in the case of *PIGM*-associated IGD, treatment of the affected children with the HDAC inhibitor NaBu, mediated histone re-acetylation and Sp1 recruitment. The role of Sp3 and Sp4 in *PIGM* transcription in the erythroid cells remains unknown.

It should be also noted that GC-boxes serve as binding motifs not only for Sp but for other TFs too. In gene promoter regions, Sp1 binding sites are frequently adjacent or can overlap with GC-binding sites for the transcription factor Egr1 (236). In megakaryocytes for example, Sp1 and Egr1 can differentially regulate transcription through the same DNA binding site (237). The role of Egr1 has been also described in cell differentiation along the monocyte lineage and interestingly, in neuronal plasticity. ChIP-seq data suggests that Egr1 is a potential candidate of *PIGM* transcription (Fig. 28).

Regarding the T cells, studies have shown the importance of GATA-3 not only in their development but also survival (155). As GATA-3 shares the same DNA-binding domain with GATA-1, is possible that GATA-3 activates *PIGM* transcription through the identified GATA sites in T cells.

The use of induced pluripotent stem (iPS) cells to produce cells from different tissue sources, has constituted a relevant tool when patient samples are not available. iPS cells could be used to generate neuronal or haematopoietic cells (in larger amounts) harbouring the -270C>G mutation, in which the regulatory mechanism could be studied in more detail, therefore contributing for the more comprehensive knowledge of the disease.

The importance of studying *cis*-regulatory mutations in the appropriate cell-type

Mutations in a Sp1 binding site, differentially affecting gene expression according to the cell type have not been reported. This may be related with the fact that most studies focus in a particular cell-type, in which there is an obvious link between the regulatory mutation and impact in gene expression. As a consequence, results might have been disregarded.

Here we showed that disruption of Sp1 binding affects *PIGM* transcription and consequently, GPI expression in a cell-type specific manner. Moreover, preliminary studies suggested that disruption of Sp1 binding might have a

functional effect in neighbouring, interacting genes that can contribute to the phenotype of the disease.

Studying the impact of a functional mutation in gene expression in a broader cell context may be particularly relevant for the accurate diagnosis of diseases sharing very similar characteristics. In this context, determining the interactions that are disrupted by the functional mutation may also provide invaluable information for the purpose of better understanding the complex regulatory mechanism behind the disease and therefore to help to design more efficient therapies.

The work presented here unveils new questions that remain to be answered. For instance, it is not known whether *PIGM*-associated IGD patients present somatic mutations that could contribute for the deficient phenotype or instead, to the normal expression of GPI. Additionally, dissecting *PIGM* transcriptional regulation at the HSC level and during haematopoietic lineage specification could provide valuable information in terms of TF binding *usage* and therefore contribute for the better comprehension of the distinct GPI phenotypes seen in the haematopoietic tissue.

Very importantly, the work presented in this thesis constitutes one, if not the first, report of cell-type specific binding preferences of a generic TF, providing novel insights into tissue-specific transcriptional control of a *housekeeping* gene.
Remarkably, these results can contribute for the better understanding of diseases caused by single point mutations that abrogate TFBS and additionally, they shed light on the importance of studying regulatory mutations at the individual cell-type level.

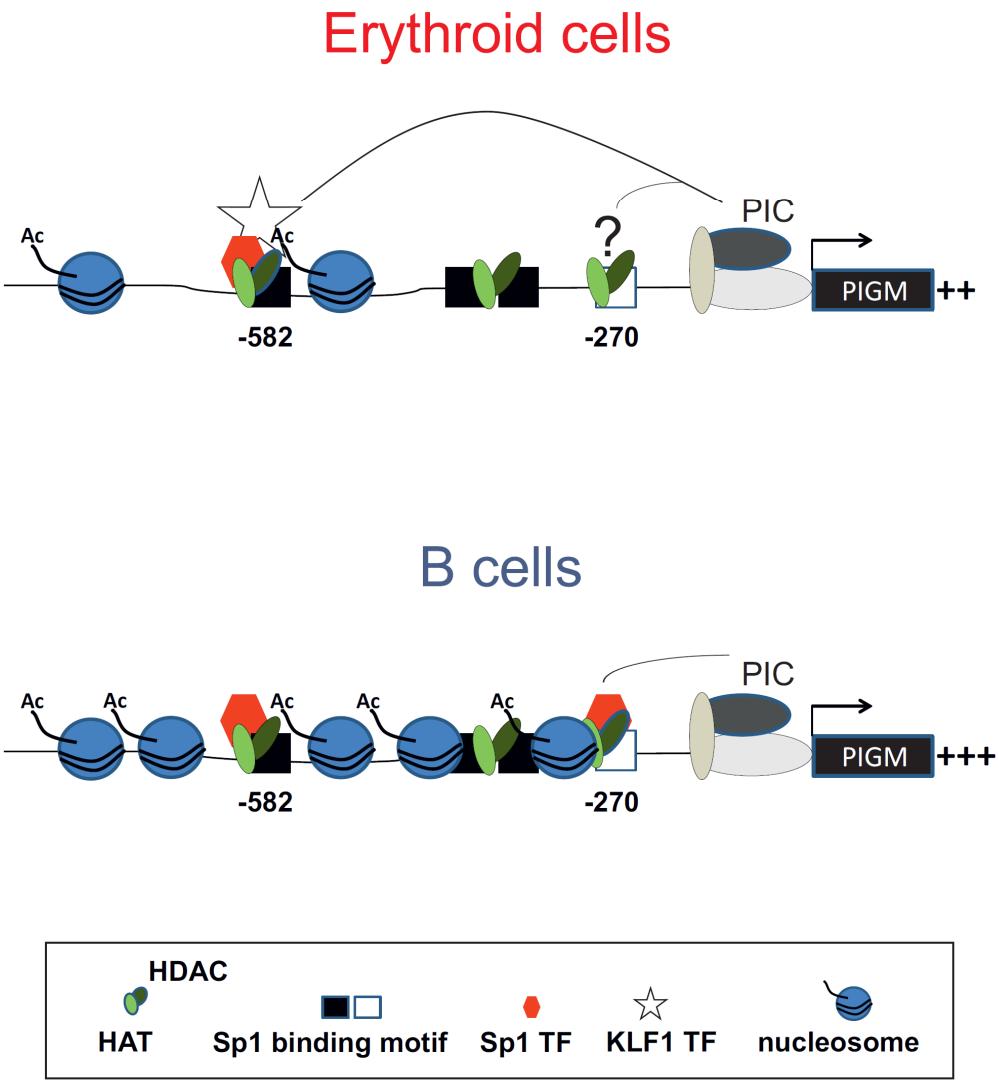


Fig. 51 – Proposed mechanism of transcriptional regulation of *PIGM* in erythroid and B cells. In WT erythroid cells, Sp1 is not recruited at the -270GC box, occupying instead the upstream GC box. In the WT B cells, Sp1 is recruited at the upstream box and also at the -270GC box. Whereas in the WT B cells, recruitment of the transcriptional machinery is mediated through Sp1-bound at the -270 GC box, in the erythroid cells, lineage-specific and generic TFs drive *PIGM* transcription. As a consequence, in the presence of the mutation, *PIGM* transcription is reduced in the B cells but it is not affected in the erythroid cells.

Appendix A

Fig. A1 – pcDNA5/FRT vector used in the stable luciferase reporter assays. The pcDNA5/FRT vector lacks the *CMV* promoter and contains a GFP reporter gene downstream at the *PIGM* promoter. -910bp of the WT and MUT *PIGM* promoter were cloned between KpnI and NcoI restriction sites.

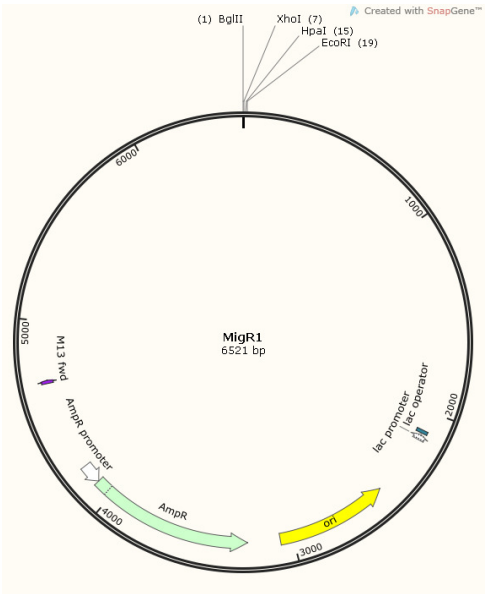


Fig. A2 – MigR1 vector used for overexpression. Sp1-mutant cDNA fragment was cloned between BglII and EcoRI in the MigR1-GFP backbone. hGATA-1 cDNA fragment was cloned in the EcoRI site in the MigR1-ratCD2 vector backbone.

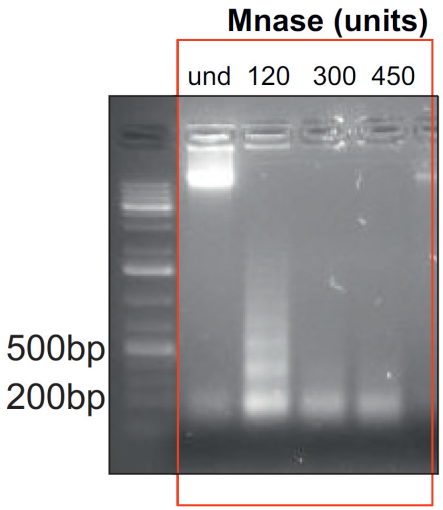


Fig. A3 – Mononuclease digestion at the indicated units monitored by gel electrophoresis (2% agarose gel).

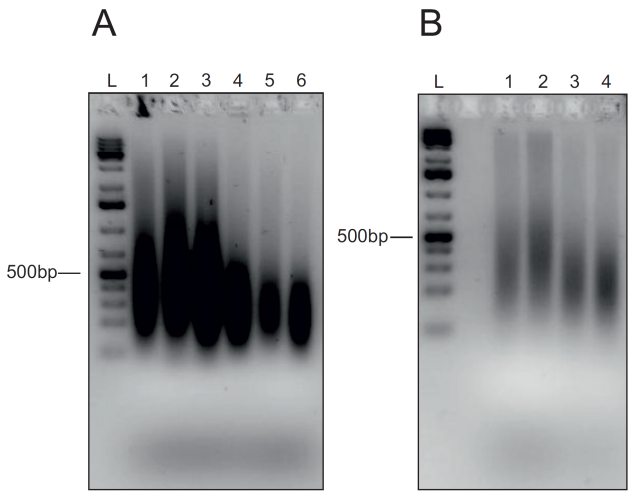


Fig. A4 – Optimization of the sonication conditions. A – Chromatin from WT LBCL (N2) sonicated for 15(1), 20(2), 22(3) and 25(4) min and chromatin from K562 sonicated for 22(5) and 25(6) min. Fragments are below 500bp. **B -** Chromatin from $CD36^-$ cells sonicated for 15(1), 20(2), 25(3) and 27(3) minutes. A period of 27 min was used for subsequent sonication of $CD36^+$ cells.

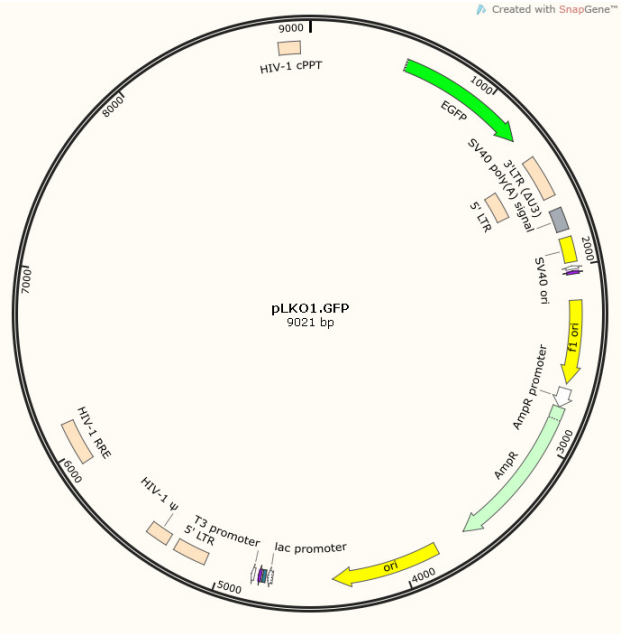


Fig. A5 – pLKO1-modified GFP vector used in the lentivirus transductions. shRNA oligonucleotides were cloned between EcoRI and AgeI restriction sites.

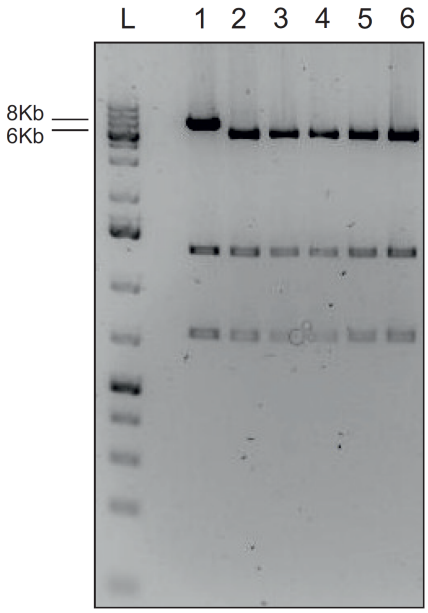


Fig. A6 – Digestion of the pLKO.1 shRNA plasmids. Approximately 100ηg of each plasmid were digested with EcoRI and NcoI. Digestion was monitored in a 2% agarose

gel. L – ladder; lane 1 – pLKO.1-GFP empty (control), lanes 2 – 6 pLKO.1-shRNA. The pLKO.1 empty vector contains a 1.9kb stuffer that is released upon digestion.

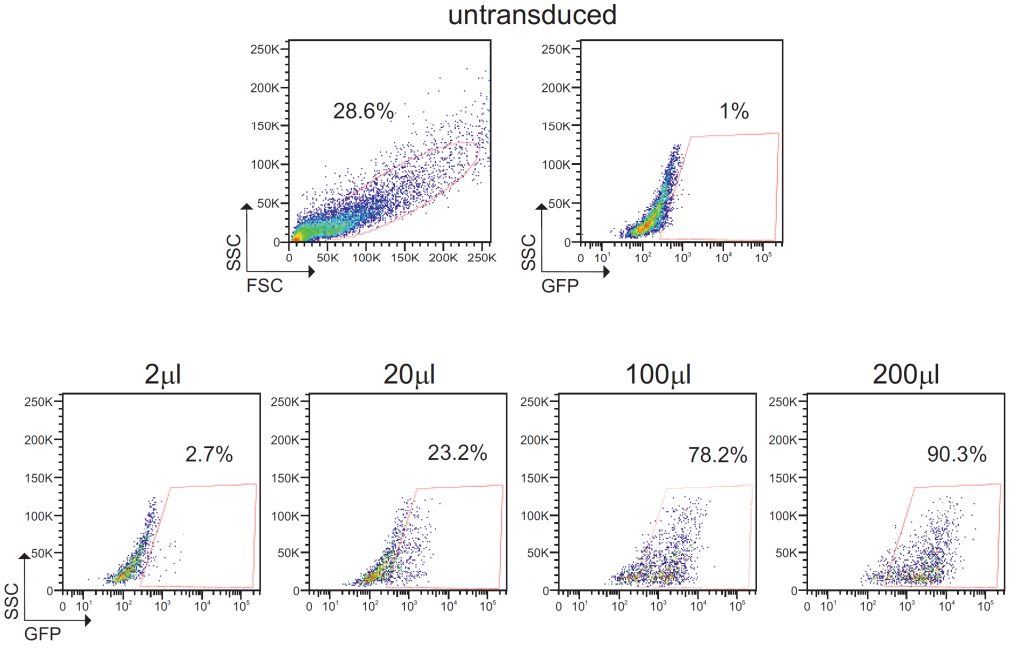


Fig. A7 – Virus titration in 293T cells. Cells were transduced with 2µl, 20µl, 100µl and 200µl of viral supernatant and then assayed for GFP expression by FACS. Viral titre (transducing units per ml) was determined according to the formula described in (238): TU ml^{-1} = (F*N*D*1000)/V, where F = percentage of GFP fluorescent cells, N = number of cells at the time of transduction (100.000), D = fold dilution of vector sample used for transduction and V = volume (µl) of diluted vector sample added into each well for transduction. The average titer was then calculated.

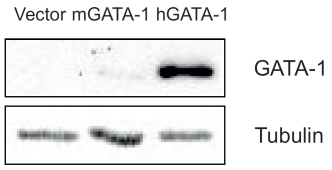


Fig. A8 – Western-Blot showing exogenous hGATA-1 expression in HeLa cells, 48 hours post-transfection. GATA-1 from murine was also tested in parallel.

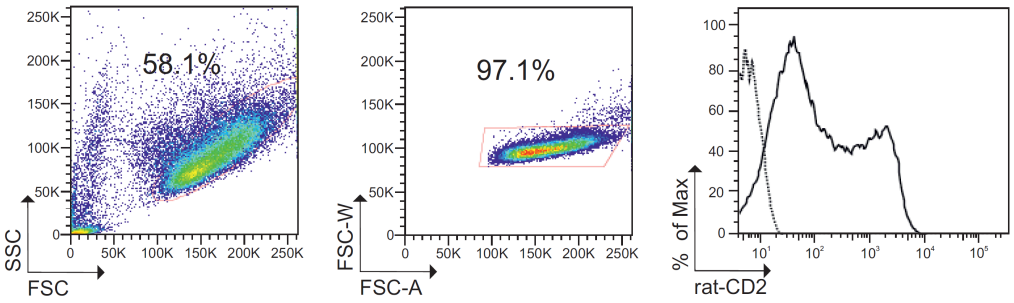


Fig. A9 – Evaluation of the transfection efficiency of the MigR1-ratCD2 plasmid in Flp-In 293 cells by FACS (solid black line).

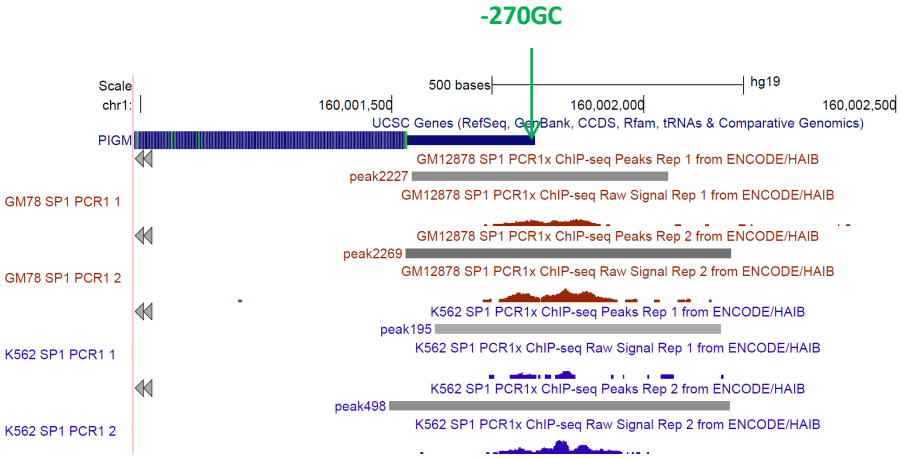


Fig. A10 – Sp1 ChIP-seq analysis at the *PIGM* promoter retrieved from the UCSC genome browser on Human Feb.2009 (GRCh37/hg19) Assembly - ENCODE database. Sp1 binding was captured in a LBCL (GM12878) and K562. The -270GC box is highlighted in green.

Appendix B – Inherited GPI Deficiency

Below are described the main genetic and clinical aspects of the group of rare inherited autosomal diseases of the GPI pathway.

PIG-A

As previously described, somatic mutations in *PIG-A* result in PNH. Although it has been suggested that germline mutations in *PIG-A* are lethal (239, 240), a recently identified X-linked disorder contradicts this supposition. Patients with a nonsense mutation on *PIG-A* (c.1234C>T) have a reduction, but not absence, of GPI biosynthesis and present anomalies such as cleft palate, neonatal seizures, contractures and central nervous system (CNS) malformations (104).

PIG-L

Compound heterozygous mutations in *PIG-L*, the enzyme that catalyses the de-N-acetylation of GlcNAc-PI in the cytoplasmic side of the ER, have been associated with the multisystemic disorder CHIME (coloboma, heart defects, icthyosis, mental insufficiency and ear defects) syndrome. The heterozygous missense mutation c.500T>C (p.Leu167Pro), located in the catalytic domain of the enzyme, was found in all six previously described CHIME patients. By employing FACS analysis in fibroblasts- and in EBV-transformed lymphoblastoid -cell lines derived from two patients, the authors observed a slight reduction (between 2- to 4-fold) in both CD59 cell-surface marker and FLAER, a modified pro-aerolysin that binds to GPI, when compared to normal controls. As discussed, the decrease in GPI expression suggests that the alteration mildly affects the function of the enzyme and GlcN-PI properly de-N-acetylated is flipped into the luminal side of the ER. Despite the lack of data in primary cells, the study confirms the pathogenicity of *PIG-L* mutations and reveals their impact at the multisystem level (105).

PIG-V

Homozygous or compound heterozygous mutations in *PIG-V*, the second mannosyltransferase encoding-gene of the pathway are known to cause HPMR syndrome, characterized as the name implies, by mental retardation and elevated levels of the GPI-anchored protein ALP in serum. Cell surface analysis of CD16 and FLAER in granulocytes from HPMR patients showed a remarkable decrease of GPI expression and rescue experiments performed in PIG-V deficient CHO (Chinese hamster ovary) cells, revealed that only PIG-V WT cDNA was able to restore GPI expression, as measured by CD59 and CD55 (78). More recently, the mechanism for release of ALP in HPMR has been demonstrated. Briefly, by employing pulse-chase metabolic labelling in PIGV-deficient and –rescued cells, the authors showed that the secretion of ALP functionally depends on the cleavage of the GPI attachment signal by the GPI transamidase complex. Moreover, they demonstrated that efficient cleavage is determined by the presence of the intermediate GPI containing at least one mannose residue.

Given the fact that half of the patients with HPMR are negative for PIG-V mutations, it was also suggested that defects in later genes of the GPI biosynthetic pathway, such as PIG-B, PIG-O and PIG-F, may also cause HPMR syndrome, contrary to defects in early GPI genes, such as PIG-L and PIG-W, in which immature GPI is not mannosylated (241). Indeed, a new study has recently recognized *PIG-O* as the second gene associated with HPMR (103).

PIG-N

Subsequently in the biosynthetic pathway of the GPI complex is *PIG-N*, the ethanolamine phosphate transferase 1-encoding gene. A study performed in mice by Hong Y. *et al*, has shown that the addition of the first ethanolamine (Etn-P) is not crucial for GPI anchor attachment as disruption of Pig-n only partially affects expression of the GPI-AP Thy-1 (73). On the other hand, a missense mutation in a highly conserved residue of *PIG-N* in humans underlies the drastic reduction (10 fold) of CD59 in fibroblasts from patients with *PIG-N* deficiency. Patients present with various abnormalities including developmental delay, cardiac, genitourinary and gastrointestinal defects, as well as severe neurological impairment with chorea and seizures, leading to early death (113). Despite the scarce

immunophenotypic analysis, the study demonstrates the vast implications associated with *PIG-N* deficiency and reinforces the importance of GPI-APs in development.

PIG-O

As mentioned above, *PIG-O* mutations have been reported as the genetic cause of HPMR, independently of mutations in *PIG-V*. Accordingly, patients with PIG-O deficiency present with neurological abnormalities (microcephaly and seizures) and elevated levels of serum ALP.

Compound heterozygous mutations, underlying GPI deficiency at the surface of granulocytes, have been identified in patients belonging to unrelated families. *PIG-O* mutations have been functionally assayed in *in vitro* experiments by transfecting two mutant PIG-O cDNAs in PIG-O-deficient cells. Whereas some *PIG-O* mutations affect protein stability, others severely impair enzyme activity reducing GPI-APs at the cell surface (103, 242).

PIGT

Very recently, mutations in *PIGT*, one of the subunits of the transamidase complex, have also been described. The homozygous mutation c.547A>C in *PIGT* is the causative defect of a previously unidentified syndrome characterised by distinct facial features, intellectual disability, hypotonia and seizures, in combination with abnormal skeletal, endocrine and ophthalmologic findings. Cell surface analysis of CD16 and FLAER in granulocytes and NK cells as well as CD59 in monocytes, revealed decreased expression of these GPI-APs in patients when compared with normal controls. Additional studies using morpholinos (MO) mediated suppression of pigt in zebrafish embryos, showed defects at the gastrulation level and in rescue experiments the presence of PIG-T homologous mRNA, but not mutated PIGT mRNA, rescued the abnormal phenotype.

Together, the results demonstrate for the first time an association between mutations in *PIGT* and human disease (114).

Recently, mutations in *PIGT* have been associated with PNH in the absence of mutations in *PIGA*. A deleterious splice site mutation simultaneously with a somatic mutation in a myeloid stem cell on *PIGT* reduces the expression of GPI in

the haematopoietic cells and causes haemolytic anaemia. The data suggests that defects in transferring intact, fully assembled GPI anchor to proteins causes PNH (115).

Bibliography

1. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. Nature genetics. 2002;32(4):623-6.

2. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell. 2002;10(6):1453-65.

3. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. The EMBO journal. 2007;26(8):2041-51.

4. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306-11.

5. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nature genetics. 2006;38(11):1348-54.

6. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome research. 2006;16(10):1299-309.

7. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013;502(7469):59-64.

8. Epstein DJ. Cis-regulatory mutations in human disease. Briefings in functional genomics & proteomics. 2009;8(4):310-6.

9. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5.

10. Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & development. 2002;16(20):2583-92.

11. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annual review of genomics and human genetics. 2006;7:29-59.

12. Malik S, Roeder RG. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nature reviews Genetics. 2010;11(11):761-72.

13. Yamashita R, Suzuki Y, Sugano S, Nakai K. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. Gene. 2005;350(2):129-36.

14. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nature genetics. 2006;38(6):626-35.

15. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(5):1412-7.

16. Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. Genomics. 1992;13(4):1095-107.

17. Zhu J, He F, Hu S, Yu J. On the nature of human housekeeping genes. Trends in genetics : TIG. 2008;24(10):481-4.

18. Lee-Huang S, Lin JJ, Kung HF, Huang PL, Lee L. The human erythropoietin-encoding gene contains a CAAT box, TATA boxes and other transcriptional regulatory elements in its 5' flanking region. Gene. 1993;128(2):227-36.

19. Yin H, Blanchard KL. DNA methylation represses the expression of the human erythropoietin gene by two different mechanisms. Blood. 2000;95(1):111-9.

20. Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, et al. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome biology. 2006;7(8):R78.

21. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature reviews Genetics. 2012;13(4):233-45.

22. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome research. 2006;16(1):55-65.

23. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, et al. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science. 2006;312(5777):1215-7.

24. Struhl K, Segal E. Determinants of nucleosome positioning. Nature structural & molecular biology. 2013;20(3):267-73.

25. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693-705.

26. Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, et al. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell. 2009;138(5):1019-31.

27. Bird A. DNA methylation patterns and epigenetic memory. Genes & development. 2002;16(1):6-21.

28. Farnham PJ. Insights from genomic profiling of transcription factors. Nature reviews Genetics. 2009;10(9):605-16.

29. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82.

30. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature. 2012;489(7414):83-90.

31. Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell. 2013;154(4):888-903.

32. Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, et al. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. Molecular cell. 2006;24(4):593-602.

33. Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. Genome research. 2008;18(11):1763-77.

34. Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. Cell. 2013;152(1-2):327-39.

35. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91-100.

36. Guertin MJ, Lis JT. Chromatin landscape dictates HSF binding to target DNA elements. PLoS genetics. 2010;6(9):e1001114.

37. John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature genetics. 2011;43(3):264-8.

38. Caputo VS, Costa JR, Makarona K, Georgiou E, Layton DM, Roberts I, et al. Mechanism of Polycomb recruitment to CpG islands revealed by inherited disease-associated mutation. Human molecular genetics. 2013.

39. Arvey A, Agius P, Noble WS, Leslie C. Sequence and chromatin determinants of cell-type-specific transcription factor binding. Genome research. 2012;22(9):1723-34.

40. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Molecular cell. 2013;52(1):25-36.

41. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome research. 2012;22(9):1658-67.

42. Demichelis F, Setlur SR, Banerjee S, Chakravarty D, Chen JY, Chen CX, et al. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(17):6686-91.

43. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013;339(6122):957-9.

44. Hashimoto K, Durbin JE, Zhou W, Collins RD, Ho SB, Kolls JK, et al. Respiratory syncytial virus infection in the absence of STAT 1 results in airway dysfunction, airway mucus, and augmented IL-17 levels. The Journal of allergy and clinical immunology. 2005;116(3):550-7.

45. Dideberg V, Kristjansdottir G, Milani L, Libioulle C, Sigurdsson S, Louis E, et al. An insertion-deletion polymorphism in the interferon regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases. Human molecular genetics. 2007;16(24):3008-16.

46. Xin M, Davis CA, Molkentin JD, Lien CL, Duncan SA, Richardson JA, et al. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(30):11189-94.

47. Shoubridge C, Tan MH, Seiboth G, Gecz J. ARX homeodomain mutations abolish DNA binding and lead to a loss of transcriptional repression. Human molecular genetics. 2012;21(7):1639-47.

48. Tischfield MA, Bosley TM, Salih MA, Alorainy IA, Sener EC, Nester MJ, et al. Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. Nature genetics. 2005;37(10):1035-7.

49. Zweier M, Gregor A, Zweier C, Engels H, Sticht H, Wohlleber E, et al. Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. Human mutation. 2010;31(6):722-33.

50. Tsai SF, Strauss E, Orkin SH. Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. Genes & development. 1991;5(6):919-31.

51. Huang DY, Kuo YY, Chang ZF. GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. Nucleic acids research. 2005;33(16):5331-42.

52. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-51.

53. Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, et al. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nature genetics. 2009;41(3):359-64.

54. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. New York: Garland Science; 2002.

55. Paulick MG, Bertozzi CR. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. Biochemistry. 2008;47(27):6991-7000.

56. Mayor S, Riezman H. Sorting GPI-anchored proteins. Nature reviews Molecular cell biology. 2004;5(2):110-20.

57. Nozaki M, Ohishi K, Yamada N, Kinoshita T, Nagy A, Takeda J. Developmental abnormalities of glycosylphosphatidylinositol-anchor-deficient embryos revealed by Cre/loxP system. Laboratory investigation; a journal of technical methods and pathology. 1999;79(3):293-9.

58. Takahashi M, Takeda J, Hirose S, Hyman R, Inoue N, Miyata T, et al. Deficient biosynthesis of N-acetylglucosaminyl-phosphatidylinositol, the first intermediate of glycosyl phosphatidylinositol anchor biosynthesis, in cell lines established from patients with paroxysmal nocturnal hemoglobinuria. The Journal of experimental medicine. 1993;177(2):517-21.

59. Miyata T, Takeda J, Iida Y, Yamada N, Inoue N, Takahashi M, et al. The cloning of PIG-A, a component in the early step of GPI-anchor biosynthesis. Science. 1993;259(5099):1318-20.

60. Takeda J, Miyata T, Kawagoe K, Iida Y, Endo Y, Fujita T, et al. Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. Cell. 1993;73(4):703-11.

61. Watanabe R, Inoue N, Westfall B, Taron CH, Orlean P, Takeda J, et al. The first step of glycosylphosphatidylinositol biosynthesis is mediated by a complex of PIG-A, PIG-H, PIG-C and GPI1. The EMBO journal. 1998;17(4):877-85.

62. Watanabe R, Kinoshita T, Masaki R, Yamamoto A, Takeda J, Inoue N. PIG-A and PIG-H, which participate in glycosylphosphatidylinositol anchor biosynthesis, form a protein complex in the endoplasmic reticulum. The Journal of biological chemistry. 1996;271(43):26868-75.

63. Hong Y, Ohishi K, Watanabe R, Endo Y, Maeda Y, Kinoshita T. GPI1 stabilizes an enzyme essential in the first step of glycosylphosphatidylinositol biosynthesis. The Journal of biological chemistry. 1999;274(26):18582-8.

64. Murakami Y, Siripanyaphinyo U, Hong Y, Tashima Y, Maeda Y, Kinoshita T. The initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-Y, a seventh component. Molecular biology of the cell. 2005;16(11):5236-46.

65. Watanabe R, Murakami Y, Marmor MD, Inoue N, Maeda Y, Hino J, et al. Initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-P and is regulated by DPM2. The EMBO journal. 2000;19(16):4402-11.

66. Watanabe R, Ohishi K, Maeda Y, Nakamura N, Kinoshita T. Mammalian PIG-L and its yeast homologue Gpi12p are N-acetylglucosaminylphosphatidylinositol de-N-acetylases essential in glycosylphosphatidylinositol biosynthesis. The Biochemical journal. 1999;339 (Pt 1):185-92.

67. Murakami Y, Siripanyapinyo U, Hong Y, Kang JY, Ishihara S, Nakakuma H, et al. PIG-W is critical for inositol acylation but not for flipping of glycosylphosphatidylinositol-anchor. Molecular biology of the cell. 2003;14(10):4285-95.

68. Maeda Y, Watanabe R, Harris CL, Hong Y, Ohishi K, Kinoshita K, et al. PIG-M transfers the first mannose to glycosylphosphatidylinositol on the lumenal side of the ER. The EMBO journal. 2001;20(1-2):250-61.

69. Ashida H, Hong Y, Murakami Y, Shishioh N, Sugimoto N, Kim YU, et al. Mammalian PIG-X and yeast Pbn1p are the essential components of

glycosylphosphatidylinositol-mannosyltransferase I. Molecular biology of the cell. 2005;16(3):1439-48.

70. Taron BW, Colussi PA, Wiedman JM, Orlean P, Taron CH. Human Smp3p adds a fourth mannose to yeast and human glycosylphosphatidylinositol precursors in vivo. The Journal of biological chemistry. 2004;279(34):36083-92.

71. Kang JY, Hong Y, Ashida H, Shishioh N, Murakami Y, Morita YS, et al. PIG-V involved in transferring the second mannose in glycosylphosphatidylinositol. The Journal of biological chemistry. 2005;280(10):9489-97.

72. Takahashi M, Inoue N, Ohishi K, Maeda Y, Nakamura N, Endo Y, et al. PIG-B, a membrane protein of the endoplasmic reticulum with a large lumenal domain, is involved in transferring the third mannose of the GPI anchor. The EMBO journal. 1996;15(16):4254-61.

73. Hong Y, Maeda Y, Watanabe R, Ohishi K, Mishkind M, Riezman H, et al. Pig-n, a mammalian homologue of yeast Mcd4p, is involved in transferring phosphoethanolamine to the first mannose of the glycosylphosphatidylinositol. The Journal of biological chemistry. 1999;274(49):35099-106.

74. Hong Y, Maeda Y, Watanabe R, Inoue N, Ohishi K, Kinoshita T. Requirement of PIG-F and PIG-O for transferring phosphoethanolamine to the third mannose in glycosylphosphatidylinositol. The Journal of biological chemistry. 2000;275(27):20911-9.

75. Ohishi K, Inoue N, Kinoshita T. PIG-S and PIG-T, essential for GPI anchor attachment to proteins, form a complex with GAA1 and GPI8. The EMBO journal. 2001;20(15):4088-98.

76. Hong Y, Ohishi K, Kang JY, Tanaka S, Inoue N, Nishimura J, et al. Human PIG-U and yeast Cdc91p are the fifth subunit of GPI transamidase that attaches GPI-anchors to proteins. Molecular biology of the cell. 2003;14(5):1780-9.

77. Ohishi K, Nagamune K, Maeda Y, Kinoshita T. Two subunits of glycosylphosphatidylinositol transamidase, GPI8 and PIG-T, form a functionally important intermolecular disulfide bridge. The Journal of biological chemistry. 2003;278(16):13959-67.

78. Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, et al. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. Nature genetics. 2010;42(10):827-9.

79. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(45):19096-101.

80. de Vooght KM, van Wijk R, van Solinge WW. Management of gene promoter mutations in molecular diagnostics. Clinical chemistry. 2009;55(4):698-708.

81. Crosby WH. Paroxysmal nocturnal hemoglobinuria; a classic description by Paul Strubling in 1882, and a bibliography of the disease. Blood. 1951;6(3):270-84.

82. Bessler M, Mason PJ, Hillmen P, Miyata T, Yamada N, Takeda J, et al. Paroxysmal nocturnal haemoglobinuria (PNH) is caused by somatic mutations in the PIG-A gene. The EMBO journal. 1994;13(1):110-7.

83. Johnson RJ, Hillmen P. Paroxysmal nocturnal haemoglobinuria: nature's gene therapy? Molecular pathology : MP. 2002;55(3):145-52.

84. Parker C, Omine M, Richards S, Nishimura J, Bessler M, Ware R, et al. Diagnosis and management of paroxysmal nocturnal hemoglobinuria. Blood. 2005;106(12):3699-709.

85. Karadimitris A, Manavalan JS, Thaler HT, Notaro R, Araten DJ, Nafa K, et al. Abnormal T-cell repertoire is consistent with immune process underlying the pathogenesis of paroxysmal nocturnal hemoglobinuria. Blood. 2000;96(7):2613-20.

86. Gargiulo L, Papaioannou M, Sica M, Talini G, Chaidos A, Richichi B, et al. Glycosylphosphatidylinositol-specific, CD1d-restricted T cells in paroxysmal nocturnal hemoglobinuria. Blood. 2013;121(14):2753-61.

87. Almeida AM, Murakami Y, Layton DM, Hillmen P, Sellick GS, Maeda Y, et al. Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. Nature medicine. 2006;12(7):846-51.

88. Suske G. The Sp-family of transcription factors. Gene. 1999;238(2):291-300.

89. Davie JR, He S, Li L, Sekhavat A, Espino P, Drobic B, et al. Nuclear organization and chromatin dynamics--Sp1, Sp3 and histone deacetylases. Advances in enzyme regulation. 2008;48:189-208.

90. Lu S, Archer MC. Sp1 coordinately regulates de novo lipogenesis and proliferation in cancer cells. International journal of cancer Journal international du cancer. 2010;126(2):416-25.

91. Knight SW, Vulliamy TJ, Morgan B, Devriendt K, Mason PJ, Dokal I. Identification of novel DKC1 mutations in patients with dyskeratosis congenita: implications for pathophysiology and diagnosis. Human genetics. 2001;108(4):299-303.

92. Salowsky R, Heiss NS, Benner A, Wittig R, Poustka A. Basal transcription activity of the dyskeratosis congenita gene is mediated by Sp1 and Sp3 and a patient mutation in a Sp1 binding site is associated with decreased promoter activity. Gene. 2002;293(1-2):9-19.

93. Di Pierro E, Cappellini MD, Mazzucchelli R, Moriondo V, Mologni D, Zanone Poma B, et al. A point mutation affecting an SP1 binding site in the promoter of the ferrochelatase gene impairs gene transcription and causes erythropoietic protoporphyria. Experimental hematology. 2005;33(5):584-91.

94. Carew JA, Pollak ES, High KA, Bauer KA. Severe factor VII deficiency due to a mutation disrupting an Sp1 binding site in the factor VII promoter. Blood. 1998;92(5):1639-45.

95. Emili A, Greenblatt J, Ingles CJ. Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. Molecular and cellular biology. 1994;14(3):1582-93.

96. Torigoe T, Izumi H, Yoshida Y, Ishiguchi H, Okamoto T, Itoh H, et al. Low pH enhances Sp1 DNA binding activity and interaction with TBP. Nucleic acids research. 2003;31(15):4523-30.

97. Soutoglou E, Viollet B, Vaxillaire M, Yaniv M, Pontoglio M, Talianidis I. Transcription factor-dependent regulation of CBP and P/CAF histone acetyltransferase activity. The EMBO journal. 2001;20(8):1984-92.

98. Sun HJ, Xu X, Wang XL, Wei L, Li F, Lu J, et al. Transcription factors Ets2 and Sp1 act synergistically with histone acetyltransferase p300 in activating human interleukin-12 p40 promoter. Acta biochimica et biophysica Sinica. 2006;38(3):194-200.

99. Doetzlhofer A, Rotheneder H, Lagger G, Koranda M, Kurtev V, Brosch G, et al. Histone deacetylase 1 can repress transcription by binding to Sp1. Molecular and cellular biology. 1999;19(8):5504-11.

100. Zhang Y, Dufau ML. Silencing of transcription of the human luteinizing hormone receptor gene by histone deacetylase-mSin3A complex. The Journal of biological chemistry. 2002;277(36):33431-8.

101. Nunes MJ, Milagre I, Schnekenburger M, Gama MJ, Diederich M, Rodrigues E. Sp proteins play a critical role in histone deacetylase inhibitor-mediated derepression of CYP46A1 gene transcription. Journal of neurochemistry. 2010;113(2):418-31.

102. Almeida AM, Murakami Y, Baker A, Maeda Y, Roberts IA, Kinoshita T, et al. Targeted therapy for inherited GPI deficiency. The New England journal of medicine. 2007;356(16):1641-7.

103. Krawitz PM, Murakami Y, Hecht J, Kruger U, Holder SE, Mortier GR, et al. Mutations in PIGO, a member of the GPI-anchor-synthesis pathway, cause hyperphosphatasia with mental retardation. American journal of human genetics. 2012;91(1):146-51.

104. Johnston JJ, Gropman AL, Sapp JC, Teer JK, Martin JM, Liu CF, et al. The phenotype of a germline mutation in PIGA: the gene somatically mutated in paroxysmal nocturnal hemoglobinuria. American journal of human genetics. 2012;90(2):295-300.

105. Ng BG, Hackmann K, Jones MA, Eroshkin AM, He P, Wiliams R, et al. Mutations in the glycosylphosphatidylinositol gene PIGL cause CHIME syndrome. American journal of human genetics. 2012;90(4):685-8.

106. Shashi V, Zunich J, Kelly TE, Fryburg JS. Neuroectodermal (CHIME) syndrome: an additional case with long term follow up of all reported cases. Journal of medical genetics. 1995;32(6):465-9.

107. Schnur RE, Greenbaum BH, Heymann WR, Christensen K, Buck AS, Reid CS. Acute lymphoblastic leukemia in a child with the CHIME neuroectodermal dysplasia syndrome. American journal of medical genetics. 1997;72(1):24-9.

108. Sidbury R, Paller AS. What syndrome is this? CHIME syndrome. Pediatric dermatology. 2001;18(3):252-4.

109. Tinschert S, Anton-Lamprecht I, Albrecht-Nebe H, Audring H. Zunich neuroectodermal syndrome: migratory ichthyosiform dermatosis, colobomas, and other abnormalities. Pediatric dermatology. 1996;13(5):363-71.

110. Rabe P, Haverkamp F, Emons D, Rosskamp R, Zerres K, Passarge E. Syndrome of developmental retardation, facial and skeletal anomalies, and hyperphosphatasia in two sisters: nosology and genetics of the Coffin-Siris syndrome. American journal of medical genetics. 1991;41(3):350-4.

111. Marcelis CL, Rieu P, Beemer F, Brunner HG. Severe mental retardation, epilepsy, anal anomalies, and distal phalangeal hypoplasia in siblings. Clinical dysmorphology. 2007;16(2):73-6.

112. Thompson MD, Nezarati MM, Gillessen-Kaesbach G, Meinecke P, Mendoza-Londono R, Mornet E, et al. Hyperphosphatasia with seizures, neurologic deficit, and characteristic facial features: Five new patients with Mabry syndrome. American journal of medical genetics Part A. 2010;152A(7):1661-9.

113. Maydan G, Noyman I, Har-Zahav A, Neriah ZB, Pasmanik-Chor M, Yeheskel A, et al. Multiple congenital anomalies-hypotonia-seizures syndrome is caused by a mutation in PIGN. Journal of medical genetics. 2011;48(6):383-9.

114. Kvarnung M, Nilsson D, Lindstrand A, Korenke GC, Chiang SC, Blennow E, et al. A novel intellectual disability syndrome caused by GPI anchor deficiency due to homozygous mutations in PIGT. Journal of medical genetics. 2013.

115. Krawitz PM, Hochsmann B, Murakami Y, Teubner B, Kruger U, Klopocki E, et al. A case of paroxysmal nocturnal hemoglobinuria caused by a germline mutation and a somatic mutation in PIGT. Blood. 2013;122(7):1312-5.

116. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes & development. 2011;25(10):1010-22.

117. Blackledge NP, Klose R. CpG island chromatin: a platform for gene regulation. Epigenetics : official journal of the DNA Methylation Society. 2011;6(2):147-52.

118. Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, et al. Role of histone H2A ubiquitination in Polycomb silencing. Nature. 2004;431(7010):873-8.

119. Landolin JM, Johnson DS, Trinklein ND, Aldred SF, Medina C, Shulha H, et al. Sequence features that drive human promoter function and tissue specificity. Genome research. 2010;20(7):890-8.

120. Supp DM, Witte DP, Branford WW, Smith EP, Potter SS. Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Developmental biology. 1996;176(2):284-99.

121. Hagen G, Muller S, Beato M, Suske G. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic acids research. 1992;20(21):5519-25.

122. Zhou X, Long JM, Geyer MA, Masliah E, Kelsoe JR, Wynshaw-Boris A, et al. Reduced expression of the Sp4 gene in mice causes deficits in sensorimotor gating and memory associated with hippocampal vacuolization. Molecular psychiatry. 2005;10(4):393-406.

123. Pinacho R, Villalmanzo N, Lalonde J, Haro JM, Meana JJ, Gill G, et al. The transcription factor SP4 is reduced in postmortem cerebellum of bipolar disorder subjects: control by depolarization and lithium. Bipolar disorders. 2011;13(5-6):474-85.

124. Pinacho R, Villalmanzo N, Roca M, Iniesta R, Monje A, Haro JM, et al. Analysis of Sp transcription factors in the postmortem brain of chronic schizophrenia: a pilot study of relationship to negative symptoms. Journal of psychiatric research. 2013;47(7):926-34.

125. Fuste M, Pinacho R, Melendez-Perez I, Villalmanzo N, Villalta-Gil V, Haro JM, et al. Reduced expression of SP1 and SP4 transcription factors in peripheral blood mononuclear cells in first-episode psychosis. Journal of psychiatric research. 2013.

126. Marin M, Karis A, Visser P, Grosveld F, Philipsen S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell. 1997;89(4):619-28.

127. Bouwman P, Gollner H, Elsasser HP, Eckhoff G, Karis A, Grosveld F, et al. Transcription factor Sp3 is essential for post-natal survival and late tooth development. The EMBO journal. 2000;19(4):655-61.

128. Terrados G, Finkernagel F, Stielow B, Sadic D, Neubert J, Herdt O, et al. Genome-wide localization and expression profiling establish Sp2 as a sequence-specific transcription factor regulating vitally important genes. Nucleic acids research. 2012;40(16):7844-57.

129. Deshane J, Kim J, Bolisetty S, Hock TD, Hill-Kapturczak N, Agarwal A. Sp1 regulates chromatin looping between an intronic enhancer and distal promoter of the human heme oxygenase-1 gene in renal cells. The Journal of biological chemistry. 2010;285(22):16476-86.

130. Opitz OG, Rustgi AK. Interaction between Sp1 and cell cycle regulatory proteins is important in transactivation of a differentiation-related gene. Cancer research. 2000;60(11):2825-30.

131. Kavurma MM, Khachigian LM. Sp1 inhibits proliferation and induces apoptosis in vascular smooth muscle cells by repressing p21WAF1/Cip1 transcription and cyclin D1-Cdk4-p21WAF1/Cip1 complex formation. The Journal of biological chemistry. 2003;278(35):32537-43.

132. Yuan P, Wang L, Wei D, Zhang J, Jia Z, Li Q, et al. Therapeutic inhibition of Sp1 expression in growing tumors by mithramycin a correlates directly with potent antiangiogenic effects on human pancreatic cancer. Cancer. 2007;110(12):2682-90.

133. Chu S. Transcriptional regulation by post-transcriptional modification--role of phosphorylation in Sp1 transcriptional activity. Gene. 2012;508(1):1-8.

134. Fischer KD, Haese A, Nowock J. Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or interference. The Journal of biological chemistry. 1993;268(32):23915-23.

135. Merika M, Orkin SH. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. Molecular and cellular biology. 1995;15(5):2437-47.

136. Sanchez HB, Yieh L, Osborne TF. Cooperation by sterol regulatory element-binding protein and Sp1 in sterol regulation of low density lipoprotein receptor gene. The Journal of biological chemistry. 1995;270(3):1161-9.

137. Sakai T, Ohtani N, McGee TL, Robbins PD, Dryja TP. Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene. Nature. 1991;353(6339):83-6.

138. Koivisto UM, Palvimo JJ, Janne OA, Kontula K. A single-base substitution in the proximal Sp1 site of the human low density lipoprotein receptor promoter as a cause of heterozygous familial hypercholesterolemia. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(22):10526-30.

139. Lower KM, De Gobbi M, Hughes JR, Derry CJ, Ayyub H, Sloane-Stanley JA, et al. Analysis of sequence variation underlying tissue-specific transcription factor binding and gene expression. Human mutation. 2013;34(8):1140-8.

140. Hoffbrand VA CD, Tuddenham EGD, Green AR Postgraduate Haematology. 6th ed. West Sussex: Wiley-Blackwell; 2010.

141. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. Cell. 2008;132(4):631-44.

142. Shivdasani RA, Mayer EL, Orkin SH. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature. 1995;373(6513):432-4.

143. Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood. 2000;96(8):2641-8.

144. Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, Radomska HS, et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(15):8705-10.

145. Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature. 1991;349(6306):257-60.

146. Weiss MJ, Keller G, Orkin SH. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. Genes & development. 1994;8(10):1184-97.

147. Kulessa H, Frampton J, Graf T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes & development. 1995;9(10):1250-62.

148. Visvader JE, Elefanty AG, Strasser A, Adams JM. GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. The EMBO journal. 1992;11(12):4557-64.

149. Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages. Cell. 2004;117(5):663-76.

150. Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, Graf T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity. 2006;25(5):731-44.

151. Mueller BU, Pabst T, Osato M, Asou N, Johansen LM, Minden MD, et al. Heterozygous PU.1 mutations are associated with acute myeloid leukemia. Blood. 2002;100(3):998-1007.

152. Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer cell. 2012;22(2):209-21.

153. Dzierzak E, Philipsen S. Erythropoiesis: development and differentiation. Cold Spring Harbor perspectives in medicine. 2013;3(4):a011601.

154. Wong P, Hattangadi SM, Cheng AW, Frampton GM, Young RA, Lodish HF. Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. Blood. 2011;118(16):e128-38.

155. Ho IC, Pai SY. GATA-3 - not just for Th2 cells anymore. Cellular & molecular immunology. 2007;4(1):15-29.

156. Weiss MJ, Orkin SH. GATA transcription factors: key regulators of hematopoiesis. Experimental hematology. 1995;23(2):99-107.

157. Molkentin JD. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. The Journal of biological chemistry. 2000;275(50):38949-52.

158. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(22):12355-8.

159. Martin DI, Orkin SH. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes & development. 1990;4(11):1886-98.

160. Omichinski JG, Trainor C, Evans T, Gronenborn AM, Clore GM, Felsenfeld G. A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(5):1676-80.

161. Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome research. 2009;19(12):2172-84.

162. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell. 1997;90(1):109-19.

163. Cohen-Kaminsky S, Maouche-Chretien L, Vitelli L, Vinit MA, Blanchard I, Yamamoto M, et al. Chromatin immunoselection defines a TAL-1 target gene. The EMBO journal. 1998;17(17):5151-60.

164. Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. The EMBO journal. 1997;16(11):3145-57.

165. Gregory RC, Taxman DJ, Seshasayee D, Kensinger MH, Bieker JJ, Wojchowski DM. Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. Blood. 1996;87(5):1793-801.

166. Boyes J, Byfield P, Nakatani Y, Ogryzko V. Regulation of activity of the transcription factor GATA-1 by acetylation. Nature. 1998;396(6711):594-8.

167. Hung HL, Lau J, Kim AY, Weiss MJ, Blobel GA. CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. Molecular and cellular biology. 1999;19(5):3496-505.

168. Imanishi M, Imamura C, Higashi C, Yan W, Negi S, Futaki S, et al. Zinc finger-zinc finger interaction between the transcription factors, GATA-1 and Sp1. Biochemical and biophysical research communications. 2010;400(4):625-30.

169. Sankaran VG, Ghazvinian R, Do R, Thiru P, Vergilio JA, Beggs AH, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. The Journal of clinical investigation. 2012;122(7):2439-43.

170. Shimizu R, Engel JD, Yamamoto M. GATA1-related leukaemias. Nature reviews Cancer. 2008;8(4):279-87.

171. Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nature genetics. 2002;32(1):148-52.

172. Feng WC, Southwood CM, Bieker JJ. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. The Journal of biological chemistry. 1994;269(2):1493-500.

173. Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. Nature genetics. 2010;42(9):801-5.

174. Arnaud L, Saison C, Helias V, Lucien N, Steschenko D, Giarratana MC, et al. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. American journal of human genetics. 2010;87(5):721-7.

175. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature. 1995;375(6529):316-8.

176. Perkins AC, Sharpe AH, Orkin SH. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature. 1995;375(6529):318-22.

177. Kaczynski J, Cook T, Urrutia R. Sp1- and Kruppel-like transcription factors. Genome biology. 2003;4(2):206.

178. Ohene-Abuakwa Y, Orfali KA, Marius C, Ball SE. Two-phase culture in Diamond Blackfan anemia: localization of erythroid defect. Blood. 2005;105(2):838-46.

179. Ronzoni L, Bonara P, Rusconi D, Frugoni C, Libani I, Cappellini MD. Erythroid differentiation and maturation from peripheral CD34+ cells in liquid culture: cellular and molecular characterization. Blood cells, molecules & diseases. 2008;40(2):148-55.

180. Brodsky RA, Mukhina GL, Li S, Nelson KL, Chiurazzi PL, Buckley JT, et al. Improved detection and characterization of paroxysmal nocturnal hemoglobinuria using fluorescent aerolysin. American journal of clinical pathology. 2000;114(3):459-66.

181. Gévry N. SA, Larochelle M., Gaudreau L. Nucleosome Mapping. In: Leblanc B, Moss T, editors. DNA-Protein Interactions2009. p. pp 281-91.

182. Eisenberg E, Levanon EY. Human housekeeping genes are compact. Trends in genetics : TIG. 2003;19(7):362-5.

183. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends in genetics : TIG. 2013;29(10):569-74.

184. Viprakasit V, Ekwattanakit S, Riolueang S, Chalaow N, Fisher C, Lower K, et al. Mutations in Kruppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. Blood. 2014.

185. Max-Audit I, Eleouet JF, Romeo PH. Transcriptional regulation of the pyruvate kinase erythroid-specific promoter. The Journal of biological chemistry. 1993;268(8):5431-7.

186. Miller IJ, Bieker JJ. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Molecular and cellular biology. 1993;13(5):2776-86.

187. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nature reviews Genetics. 2009;10(10):669-80.

188. Valverde-Garduno V, Guyot B, Anguita E, Hamlett I, Porcher C, Vyas P. Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. Blood. 2004;104(10):3106-16.

189. Moore CB, Guthrie EH, Huang MT, Taxman DJ. Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown. Methods Mol Biol. 2010;629:141-58.

190. Woon Kim Y, Kim S, Geun Kim C, Kim A. The distinctive roles of erythroid specific activator GATA-1 and NF-E2 in transcription of the human fetal gamma-globin genes. Nucleic acids research. 2011;39(16):6944-55.

191. Bieker JJ. Isolation, genomic structure, and expression of human erythroid Kruppel-like factor (EKLF). DNA and cell biology. 1996;15(5):347-52.

192. Ko LJ, Engel JD. DNA-binding specificities of the GATA transcription factor family. Molecular and cellular biology. 1993;13(7):4011-22.

193. Merika M, Orkin SH. DNA-binding specificity of GATA family transcription factors. Molecular and cellular biology. 1993;13(7):3999-4010.

194. Johnson KD, Grass JA, Boyer ME, Kiekhaefer CM, Blobel GA, Weiss MJ, et al. Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(18):11760-5.

195. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(15):8811-6.

196. Im H, Grass JA, Johnson KD, Kim SI, Boyer ME, Imbalzano AN, et al. Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(47):17065-70.

197. Grass JA, Jing H, Kim SI, Martowicz ML, Pal S, Blobel GA, et al. Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. Molecular and cellular biology. 2006;26(19):7056-67.

198. Fujiwara T, O'Geen H, Keles S, Blahnik K, Linnemann AK, Kang YA, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Molecular cell. 2009;36(4):667-81.

199. Yu M, Riva L, Xie H, Schindler Y, Moran TB, Cheng Y, et al. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Molecular cell. 2009;36(4):682-95.

200. Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Molecular cell. 2005;17(3):453-62.

201. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell. 2011;147(6):1408-19.

202. Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, et al. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. The EMBO journal. 2005;24(13):2367-78.

203. Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, et al. GATA-1 forms distinct activating and repressive complexes in erythroid cells. The EMBO journal. 2005;24(13):2354-66.

204. Saleque S, Kim J, Rooke HM, Orkin SH. Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. Molecular cell. 2007;27(4):562-72.

205. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823-6.

206. Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome research. 2010;20(8):1052-63.

207. Woo AJ, Moran TB, Schindler YL, Choe SK, Langer NB, Sullivan MR, et al. Identification of ZBP-89 as a novel GATA-1-associated transcription factor involved in megakaryocytic and erythroid development. Molecular and cellular biology. 2008;28(8):2675-89.

208. Lamonica JM, Vakoc CR, Blobel GA. Acetylation of GATA-1 is required for chromatin occupancy. Blood. 2006;108(12):3736-8.

209. Yu L, Ji W, Zhang H, Renda MJ, He Y, Lin S, et al. SENP1-mediated GATA1 deSUMOylation is critical for definitive erythropoiesis. The Journal of experimental medicine. 2010;207(6):1183-95.

210. Schwarzmayr L, Andorfer P, Novy M, Rotheneder H. Regulation of the E2F-associated phosphoprotein promoter by GC-box binding proteins. The international journal of biochemistry & cell biology. 2008;40(12):2845-53.

211. Li L, He S, Sun JM, Davie JR. Gene regulation by Sp1 and Sp3. Biochemistry and cell biology = Biochimie et biologie cellulaire. 2004;82(4):460-71.

212. Chen MJ, Shimada T, Moulton AD, Cline A, Humphries RK, Maizel J, et al. The functional human dihydrofolate reductase gene. The Journal of biological chemistry. 1984;259(6):3933-43.

213. Dynan WS, Sazer S, Tjian R, Schimke RT. Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature. 1986;319(6050):246-8.

214. Al-Sarraj A, Day RM, Thiel G. Specificity of transcriptional regulation by the zinc finger transcription factors Sp1, Sp3, and Egr-1. Journal of cellular biochemistry. 2005;94(1):153-67.

215. Malek A, Nunez LE, Magistri M, Brambilla L, Jovic S, Carbone GM, et al. Modulation of the activity of Sp transcription factors by mithramycin analogues as a new strategy for treatment of metastatic prostate cancer. PloS one. 2012;7(4):e35130.

216. Blume SW, Snyder RC, Ray R, Thomas S, Koller CA, Miller DM. Mithramycin inhibits SP1 binding and selectively inhibits transcriptional activity of the dihydrofolate reductase gene in vitro and in vivo. The Journal of clinical investigation. 1991;88(5):1613-21.

217. Barcelo F, Scotta C, Ortiz-Lombardia M, Mendez C, Salas JA, Portugal J. Entropically-driven binding of mithramycin in the minor groove of C/G-rich DNA sequences. Nucleic acids research. 2007;35(7):2215-26.

218. Encarnacao PC, Ramirez VP, Zhang C, Aneskievich BJ. Sp sites contribute to basal and inducible expression of the human TNIP1 (TNFalpha-inducible protein 3-interacting protein 1) promoter. The Biochemical journal. 2013;452(3):519-29.

219. Birnbaum MJ, van Wijnen AJ, Odgren PR, Last TJ, Suske G, Stein GS, et al. Sp1 trans-activation of cell cycle regulated promoters is selectively repressed by Sp3. Biochemistry. 1995;34(50):16503-8.

220. Yu B, Datta PK, Bagchi S. Stability of the Sp3-DNA complex is promoter-specific: Sp3 efficiently competes with Sp1 for binding to promoters containing multiple Sp-sites. Nucleic acids research. 2003;31(18):5368-76.

221. Kimura N, Takamatsu N, Yaoita Y, Osamura RY. Identification of transcriptional regulatory elements in the human somatostatin receptor sst2 promoter and regions including estrogen response element half-site for estrogen activation. Journal of molecular endocrinology. 2008;40(2):75-91.

222. Ou XM, Chen K, Shih JC. Dual functions of transcription factors, transforming growth factor-beta-inducible early gene (TIEG)2 and Sp3, are mediated by CACCC

element and Sp1 sites of human monoamine oxidase (MAO) B gene. The Journal of biological chemistry. 2004;279(20):21021-8.

223. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. Cell. 2009;136(1):175-86.

224. Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell reports. 2012;2(1):198-205.

225. Marella NV, Bhattacharya S, Mukherjee L, Xu J, Berezney R. Cell type specific chromosome territory organization in the interphase nucleus of normal and cancer cells. Journal of cellular physiology. 2009;221(1):130-8.

226. Iborra FJ, Pombo A, Jackson DA, Cook PR. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. Journal of cell science. 1996;109 (Pt 6):1427-36.

227. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nature genetics. 2004;36(10):1065-71.

228. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes & development. 2012;26(1):11-24.

229. Heuser K, Nagelhus EA, Tauboll E, Indahl U, Berg PR, Lien S, et al. Variants of the genes encoding AQP4 and Kir4.1 are associated with subgroups of patients with temporal lobe epilepsy. Epilepsy research. 2010;88(1):55-64.

230. Bassi MT, Bresolin N, Tonelli A, Nazos K, Crippa F, Baschirotto C, et al. A novel mutation in the ATP1A2 gene causes alternating hemiplegia of childhood. Journal of medical genetics. 2004;41(8):621-8.

231. Kolesnikova TV, Kazarov AR, Lemieux ME, Lafleur MA, Kesari S, Kung AL, et al. Glioblastoma inhibition by cell surface immunoglobulin protein EWI-2, in vitro and in vivo. Neoplasia. 2009;11(1):77-86, 4p following

232. Calpe S, Wang N, Romero X, Berger SB, Lanyi A, Engel P, et al. The SLAM and SAP gene families control innate and adaptive immune responses. Advances in immunology. 2008;97:177-250.

233. Hasan NM, MacDonald MJ. Sp/Kruppel-like transcription factors are essential for the expression of mitochondrial glycerol phosphate dehydrogenase promoter B. Gene. 2002;296(1-2):221-34.

234. Guo H, Mi Z, Kuo PC. Characterization of short range DNA looping in endotoxin-mediated transcription of the murine inducible nitric-oxide synthase (iNOS) gene. The Journal of biological chemistry. 2008;283(37):25209-17.

235. Milagre I, Nunes MJ, Gama MJ, Silva RF, Pascussi JM, Lechner MC, et al. Transcriptional regulation of the human CYP46A1 brain-specific expression by Sp transcription factors. Journal of neurochemistry. 2008;106(2):835-49.

236. Huang RP, Fan Y, Ni Z, Mercola D, Adamson ED. Reciprocal modulation between Sp1 and Egr-1. Journal of cellular biochemistry. 1997;66(4):489-99.

237. Gannon AM, Kinsella BT. Regulation of the human thromboxane A2 receptor gene by Sp1, Egr1, NF-E2, GATA-1, and Ets-1 in megakaryocytes. Journal of lipid research. 2008;49(12):2590-604.

238. Kutner RH, Zhang XY, Reiser J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. Nature protocols. 2009;4(4):495-505.

239. Keller P, Payne JL, Tremml G, Greer PA, Gaboli M, Pandolfi PP, et al. FES-Cre targets phosphatidylinositol glycan class A (PIGA) inactivation to hematopoietic stem cells in the bone marrow. The Journal of experimental medicine. 2001;194(5):581-9.

240. Keller P, Tremml G, Rosti V, Bessler M. X inactivation and somatic cell selection rescue female mice carrying a Piga-null mutation. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(13):7479-83.

241. Murakami Y, Kanzawa N, Saito K, Krawitz PM, Mundlos S, Robinson PN, et al. Mechanism for release of alkaline phosphatase caused by glycosylphosphatidylinositol deficiency in patients with hyperphosphatasia mental retardation syndrome. The Journal of biological chemistry. 2012;287(9):6318-25.

242. Kuki I, Takahashi Y, Okazaki S, Kawawaki H, Ehara E, Inoue N, et al. Vitamin B6-responsive epilepsy due to inherited GPI deficiency. Neurology. 2013;81(16):1467-9.